

NBS
PUBLICATIONS

NAT'L INST. OF STAND & TECH
A11106 430952

NBSIR 85-3278

A Self-Evaluative Laboratory Quality System

An NBS Research Associates' Report

by

Carol J. Kelly, Kim D. Bruley
David H. Craig, & Dale J. Pangonis
Central Laboratory, Ford Motor Company

and

John W. Locke
National Bureau of Standards

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Office of Product Standards Policy
Gaithersburg, MD 20899

November 1985



DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

QC
100
U56
85-3278
1985
c. 2

NBS-CIC

OC100

.USB

NO. 85-3278

1985

C.2

NBSIR 85-3278

**A SELF-EVALUATIVE LABORATORY
QUALITY SYSTEM**

An NBS Research Associates' Report

by

Carol J. Kelly, Kim D. Bruley
David H. Craig, & Dale J. Pangonis
Central Laboratory, Ford Motor Company

and

John W. Locke
National Bureau of Standards

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Office of Product Standards Policy
Gaithersburg, MD 20899

November 1985

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
DISCLAIMER	iii
I. INTRODUCTION	1
II. DEFINITIONS	2
Precision and Accuracy	2
Quality	2
Quality Control	3
Quality Assurance	3
Quality System	4
III. MAE BASED QUALITY SYSTEM FOR LABORATORY SELF-EVALUATION	4
III.1 EXPERIENCE WITH MAEs	5
III.2 A GENERALIZED SELF-EVALUATIVE QUALITY SYSTEM	6
III.3 ESTABLISHING A SYSTEM	6
Determination of Need	6
Selection of Measurements	7
Establish Working Groups	7
Develop Measurement Assurance Plan	7
Implement MAE	8
III.4 GENERIC QUALITY SYSTEM ELEMENTS (CRITERIA)	8
Organization	8
Personnel	8
Facilities	9
Procedures	9
Reporting	9
IV. REFERENCES	11

APPENDICES

A. MEASUREMENT ASSURANCE EXPERIMENTS	13
A.1 THERMOGRAVIMETRIC ANALYSIS	13
Technique and Equipment	13
Experimental Plan	13
Results	14
Plans for Future Action	15
Conclusions	15
A.2 GAGE BLOCK TOLERANCE TESTING	18
Experimental Plan	18
Results	18
Plans for Future Action	19
Calibration of Working Masters	20
Conclusions	21
A.3 LECO CARBON ANALYSIS	23
Equipment and Operation	23
Experimental Plan	24
Results	24
Conclusions and Recommendations	26

A.4	COOLANT INHIBITORS BY HIGH PERFORMANCE LIQUID CHROMATOGRAPHY	28
	Technique and Equipment	28
	Experimental Plan	29
	Results	29
	Plans for Future Action	31
	Conclusions	31
B.	OTHER QUALITY SYSTEMS FOR SELF-EVALUATION	33
B.1	MEASUREMENT ASSURANCE PROGRAMS (MAPs)	33
B.2	LABORATORY ACCREDITATION	34
	Organization Criteria	34
	Personnel Criteria	34
	Facility and Environmental Criteria	34
	Equipment and Calibration Criteria	35
	Sample Handling Criteria	35
	Test Method and Procedure Criteria	35
	Test Report Criteria	35
	Record Criteria	35
	Quality System Criteria	36
B.3	INTERNAL QUALITY AUDIT	36
	Laboratory Organization	36
	Evaluation and Detection	37
	Assessment	37
	Feedback - The Quality Audit Report	37
	Cost versus Benefit	37
C.	QUALITY SYSTEMS IN OTHER LABORATORIES	38
C.1	OIL/CHEMICAL INDUSTRY LABORATORY	38
C.2	BUILDING MATERIALS LABORATORY	38
C.3	CHEMICAL INDUSTRY LABORATORY	39
C.4	SUBCONTRACT LABORATORIES	40

FIGURES

Figure 1:	TGA Thermogram	42
Figure 2:	Thermogravimetric Balance Assembly	43
Figure 3:	Run Chart for % Elastomer and % Carbon	44
Figure 4:	Run Chart for % Ash	45
Figure 5:	% Ash versus Sample Weight	46
Figure 6:	% Carbon versus % Elastomer	47
Figure 7:	% Ash versus % Elastomer	48
Figure 8:	Quality Comparison - English Set 1	49
Figure 9:	Quality Comparison - English Set 2	49
Figure 10:	Quality Comparison - Metric	50
Figure 11:	Average % Carbon Control Chart NBS SRM 3c	51
Figure 12:	Range % Carbon Control Chart NBS SRM 3c	51
Figure 13:	Average % Carbon Control Chart - AR302	52
Figure 14:	Range % Carbon Control Chart - AR302	52
Figure 15:	Observed % Carbon versus Certified	53
Figure 16:	High Performance Liquid Chromatography Instrumentation	54
Figure 17:	Typical HPLC Chromatogram	55
Figure 18:	Variability of TT Reference Standard	56
Figure 19:	Variability of TT Coolant Sample 1	56
Figure 20:	Variability of TT Coolant Sample 2	56
Figure 21:	Variability of MBT Coolant Sample 3	56
Figure 22:	Chromatogram Illustrating Phenol Interference	57

ACKNOWLEDGEMENTS

The Ford Research Associates are indebted to William A. Goering, Central Laboratory Manager, for his continued support, funding, encouragement, helpful advice, and patience during the planning, implementation, and documentation of this NBS Research Associate Program.

The Research Associates also thank Harvey W. Berger, Gerald A. Berman, Peter R. deBruyn, Mary G. Natrella, M. Carrol Croarkin, and John K. Taylor, of the National Bureau of Standards for helpful and informative discussions during the course of this program. Brian Belanger, Keith Eberhart, Harry Ku, Theodore Lashof, Jeffrey Horlick, Albert Tholen, George Schultz, John Simpson, John Mandel, and Clyde Tucker, all of NBS, are also thanked for providing helpful information.

The authors thank the following dedicated Central Laboratory personnel who willingly participated in the development of the MAEs.

- Thermogravimetric Analysis-Arsen Terjimanian, Linda Graham, Maria Korcek;
- Gage Block Tolerance Testing-Joe Schwartz, Walter Kuhnert;
- LECO Carbon Analysis-Dave Spaight, Kathy Berry, Curt Horvath, Tim Thompson;
- Coolant Inhibitors by High Performance Liquid Chromotography-Diago Lopes;

And James Galloway, Roger Dubie, and Fred Vitale, for helpful discussion. W. W. Scherkenback is thanked for the opportunity to interact and share with W. Edwards Deming the progress of the MAEs.

DISCLAIMER

The use of trade names and manufacturer's names in this report in no way represents an endorsement of a particular product or manufacturer by the National Bureau of Standards or Ford Motor Company.

I. INTRODUCTION

Users of laboratory services, whether they are departments of the same parent company or commercial clients, are increasingly interested in the quality of the testing and test data reported to them. Laboratory management requires system and techniques to evaluate and demonstrate the competence of laboratory operations and quality of the reported data. Those who inspect, accredit, and certify laboratories, seek improvement in laboratory assessment methods. This paper is the result of research on techniques for self-evaluation of laboratory quality carried out by the Central Laboratory of Ford Motor Company in cooperation with the National Bureau of Standards (NBS).

Basic techniques for assessing the quality of laboratory services include Measurement Assurance Programs (MAPs), laboratory accreditation, and internal quality audits. This paper describes research at Central Laboratory on Measurement Assurance Experiments (MAEs) to explore the use of within-laboratory testing to evaluate data quality. MAEs are laboratory testing protocols intended to provide appropriate data for statistical measures of precision and accuracy that a laboratory can achieve in performing a given test method or group of related methods. MAEs are an excellent way to determine the attainable precision of laboratory measurements. They can be used to explore and demonstrate how individual factors (e.g., people, equipment, methods, and samples) contribute to a measurement. Laboratory management can use the results of these experiments to demonstrate the quality of test data to clients. Laboratory accreditors, auditors, and certifiers may use the results to more efficiently assess the competence of laboratories. Performing MAEs provides laboratory staff with an improved understanding of the measurement process and its inherent uncertainties.

MAEs may be expensive and, therefore, not applicable to all test methods. The research reported in the paper results in a generalized plan for a self-evaluative quality system which combines MAEs with the criteria developed for the NBS National Voluntary Laboratory Accreditation Program (NVLAP). MAEs relate to specific measurement technologies and the criteria address overall laboratory management and practices. Individual laboratories seeking to establish such a quality system must adapt the components of the system to suit their own goals and quality policies. Central Laboratory is selectively implementing portions of the self-evaluative quality system reported here. MAEs are being developed to measure and document precision and/or accuracy, and development of a consensus quality policy has begun.

In 1983, NBS established a Research Associate Program with Central Laboratory to study laboratory quality systems. Central Laboratory's objective was to develop a self-evaluative system applicable to its own needs. NBS was interested in accreditation system, the practical applications of statistical procedures in testing laboratories, and "good laboratory practices." NBS' staff participating in the program had extensive experience in laboratory accreditation and quality assurance programs [1].

Central Laboratory is a broad-based materials testing laboratory. Test data are provided to meet company needs using a variety of test methods on different materials on an as-needed basis. The laboratory does not operate as a quality control laboratory, for which well-developed quality systems are available and for which "out-of-control" conditions (of either the manufacturing or the measurement process) are readily apparent so that corrective action can be taken quickly. The challenge for Central Laboratory has been, therefore, to develop an economical self-evaluative quality system for broad-based laboratories.

Quality, quality control, quality assurance, and quality system, are terms generally applied to manufacturing processes. In the second section of this paper, the authors present their definitions of those terms, along with those for precision and accuracy, as measures of data quality in a laboratory situation. The third section of the paper describes the generalized self-evaluative quality system developed as a result of the research associate program between Central Laboratory and NBS. The MAEs upon which the generalized system was based are described in detail in appendix A. Other quality systems for self-evaluation are presented briefly in appendix B

II. DEFINITIONS

Terminology is especially important to understanding quality systems. Juran [2] summarizes the confusion about quality terminology with the statement: "One must cope not only with multiple terms for the same concepts but also with single terms that have multiple meanings." Laboratory clients cannot make objective judgements about the quality of laboratory data without appropriate definition of data quality.

Precision and Accuracy

Precision is the statistically determined random error associated with a measurement process in control. If a measurement process is in a state of statistical control and if a measurement is repeated over a period of time, the result will vary in a statistically predictable manner. Accuracy is more difficult to ascertain than precision. The true value of anything is a concept and it can only be estimated with some measure of uncertainty [3]. Accuracy is, therefore, a measure of the closeness of the estimate to the true value and includes errors of precision and bias. (Bias is an offset or nonzero correction to the measurement process.)

Using precision and accuracy as objective measures of data quality can provide the means to match the data quality to the intended use and decide if the measurement is "good enough." Documented precision and accuracy for laboratory test data provides analysts, clients, laboratory managers, and technical peers with information to objectively evaluate data quality. It can help identify needs and set priorities for improvement of measurements.

Laboratory measurements are used to make a variety of decisions, including those about material suitability, health, safety, acceptability of performance characteristics, and conformance to product specifications. Precision and accuracy required for various decisions can, and should, differ for economic reasons. The selection of precision and accuracy levels for a given decision requires consideration of the economic consequences of measurement errors. For example, the assay of platinum and gold demands both good precision and high accuracy because of the value of precious metals. The presence of water in gasoline may be determined with low precision and accuracy since the measurement is not critical to the determination if water content will cause automobiles to stall.

Quality

"Quality is the totality of features and characteristics of a product or service that bear on its ability to satisfy given needs [4]. Quality of a manufactured product is the capability of that product to meet the user's needs in a cost-effective manner. Product features and characteristics include design, fit, finish, ease of repair, and durability or freedom from failure. The needs of the user define the quality limits. "Acceptable"

quality for one user may be "unacceptable" quality for another. A user's perception of quality also includes hard-to-measure and intangible features.

Quality features and characteristics of a laboratory include capability, oral and written communication of results, precision, accuracy, timeliness, image, accessibility, and technical credibility. Some of these quality characteristics are tangible and measurable; others are intangible and difficult or impossible to quantify, but nonetheless important to the user's perception of quality.

Quality Control

Quality control is a planned system of activities to provide a level of quality that meets the needs of users and the use of such a system. The objective of quality control is to provide a product or service that has the required features of safety, adequacy, dependability, and economy. The overall quality control system integrates many related functions including; proper specification of the product or service, production to meet the full intent of the specification, inspection to determine if the resulting material, product, or service is in compliance with the specification, and review of in-use performance to determine if revision of the specification is necessary [5].

"Quality control looks at the issue from the manufacturer's point of view, ensuring that the products going out the door meet all design specifications." [6] In manufacturing processes, quality objectives have historically been set by means of product or material specifications. The tools that are specific to quality control are: sampling, testing, measuring, inspection, and all forms of statistical process control. Quality control makes use of these tools to monitor the manufacturing process and assure that the specifications are met. The broader issue of meeting the ultimate user's needs is not considered an objective of quality control.

For a laboratory, quality control includes: conducting the proper test, proper standards and procedures, calibrated equipment, good laboratory practices, control charts, validated methods, documenting what was done for data traceability, and inspecting the report for completeness [7].

Quality Assurance

Quality assurance is a planned system of activities for continuous evaluation of the adequacy and effectiveness of a quality control program with provision for initiating corrective action where necessary. "Quality assurance is the overall set of methods and techniques used to assure the quality from design through the product's life cycle. During the quality assurance process, much emphasis is put on looking at the product from the customer's viewpoint [6].

The major tools for quality assurance are: verification, audits, warranty returns, customer feedback, market research and other means of verifying and providing evidence, to all concerned, that quality objectives for a specific material, product, service, system or environment are being met [8]. Quality assurance programs may also include interlaboratory testing programs, round-robin testing, blind knowns, peer reviews, and participation in NBS Measurement Assurance Programs (MAPs).

An example of an internal laboratory quality audit is described by Gaft and Richards [9]. This audit provides a dynamic interactive mechanism to identify system or procedural problems, make recommendations for solutions and provide

immediate feedback. It establishes confidence in the laboratory product to management, clients and others. However, it only addresses systems or procedures and cannot assure the correctness of individual test results. Sample mix-ups, deviations from procedures or the use of improper standards or procedures may not be identified. These factors must be addressed by other quality system functions.

Quality System

The combination of quality assurance and quality control constitutes a quality system. Organization of the quality assurance and the quality control subsystems in a dynamic, interdependent manner provides services or products, which are satisfactory for use, as well as evidence thereof to management, clients, peers, and others.

The implementation of a quality system is complex because of the interdependence of the tools, the objectives and the responsibilities. The evolution of quality systems in the automotive industry has been confused by the interpretation of the components of a quality system. For example, a distinction between quality control and quality assurance does not exist in Ford Motor Company's manufacturing activities. Activities are described as quality control and quality assurance, but these terms don't necessarily define the functions performed. In fact, the Ford Motor Company quality manual does not use the terms quality control or quality assurance [10]. Company management has instead defined quality objectives that encompass all activities: suppliers, purchasing, government regulations, safety, production, engineering, design, field service, marketing and sales. Cooperative interaction in meeting the quality objectives provides a dynamic frame work for never-ending quality improvements.

A laboratory quality system also interacts with internal and external systems including the parent company, government, and industry. A laboratory often serves a client who is unfamiliar with the purchased service. This is similar to other technical professions, (i.e., doctors, lawyers and accountants), where the service recipient is relying on the professional for information and decisions about unfamiliar matters. The quality of our doctors, lawyers and accountants is evaluated using professional accreditation, certification, educational background, successful experiences, published performance measures and satisfied client referrals. Laboratory clients have a similar need to evaluate laboratory quality and can use similar resources to do so.

III. MAE BASED QUALITY SYSTEM FOR LABORATORY SELF-EVALUATION

A laboratory quality system must deal with the tangible and the intangible aspects of quality. By auditing procedures and subsystems, a knowledgeable person can develop a subjective measure of the data quality that may be expected from intangible quality system elements.

Laboratory accreditation and certification procedures address the tangible aspects of quality system through proficiency testing and extensive quality assurance procedures which are applied to specific test methods. It is not acceptable to adopt the precision and accuracy statement of standards or equipment specifications because they are not usually relevant to the tests conducted by a laboratory that did not participate in their development. To be valid to a particular laboratory, statements about precision and accuracy must relate to the actual measurement produced by that laboratory. The MAEs described in this paper were developed by Central Laboratory to meet its own specific needs for quality limits of accuracy and precision.

III.1 FORD EXPERIENCE WITH MAEs

Four test methods upon which to base MAEs were selected by the authors (Central Laboratory) to meet two objectives: (1) to obtain specific statistical measures of precision, accuracy and process control, and (2) to identify the prerequisites and evaluate the benefits of MAEs. Each of the four Ford authors selected a test method and developed a plan to obtain a measure of precision and an estimate of accuracy.

Reference materials were available for the gage block tolerance testing and the LECO carbon analysis so that estimates of accuracy could be made. Standards are unavailable for thermal analysis of polymers and High Performance Liquid Chromatography (HPLC) of coolant additives therefore alternative methods had to be used to obtain an estimate of accuracy.

Implementing the four MAEs was a significant learning experience for Central Laboratory personnel and contributed to improved technical skills. Measurement variations among analysts performing the same test is probably the least controllable factor in most laboratories. MAEs are mechanisms to transfer expertise within a laboratory group among its members. For the group performing the LECO carbon analysis, this training benefit became apparent when a new analyst joined the group and generated data with a wider variation that had previously been experienced.

Use of statistical methods in performing the MAEs resulted in improved understanding of the measurement processes by Central Laboratory staff. Statistical analysis benefits the learning process because it provides feedback which is immediate, objective and scientifically based and increases the appreciation of parameters affecting the measurement process. For example, the thermal analysis MAE group gained increased confidence in the measurement they were making, but the groups working on the gage block inspection process and HPLC analysis of coolants MAEs raised valid dissatisfaction with those measurement processes.

Experience in each of the four MAEs confirmed the need to document procedures and thereby assure that each process was precisely followed by all analysts. The importance of good documentation increased with the number of analysts participating in the MAE. Even the HPLC analysis of coolants, conducted by a single analyst, produced measurement variation that could not be traced without careful documentation.

Necessary ingredients for a successful MAE are: one or more analysts having specific technical skills applicable to the test method, good interpersonal relationships among all members of the group, and an environment that encourages inquiry. The analysts' participation in planning, developing, and implementing the MAEs was critical to their success and provided confidence that the measurement procedures were followed.

Additional benefits of performing MAEs were the development of improved testing capabilities. The thermogravimetric analysis MAE gave assurance that the analysts' techniques and laboratory equipment were in statistical control. The LECO carbon analysis MAE extended precision and/or accuracy limits to a wider range of analyses. The MAE on carbon determination in cast iron (2-4% carbon) is also applicable to analyses of carbon in steel (0.01 - 1.0% carbon) because the analysts gained skill in identifying and controlling the critical elements of the measurement process. The MAE experience reduced specific causes of measurement variation: equipment malfunction, calibration procedures, sample variation, sampling procedures and other analyst variables.

A large volume of samples is important to the success of a MAE in a broad-based laboratory. A high volume of ongoing testing is particularly helpful since the additional cost for running control samples is small. It takes longer, or costs more, to obtain enough data for sound statistical evaluation if a large number of samples are not being analyzed by the laboratory on a regular basis. Additionally, it is hard to sustain analyst interest, enthusiasm or motivation to continue the MAE if testing frequency is low.

Client understanding of, and demand for, precision and accuracy of measurements is important to the success of a MAE. Clients can provide incentive to maintain continuous measurement of data quality. Client understanding and appreciation of the MAEs reported here was not apparent. In fact, attempts to explain the precision of the LECO carbon measurement were not appreciated. A workable short term substitute for client recognition may be laboratory management support for, and recognition of the benefits of, documented precision and accuracy. Until clients understand and demand information about precision and accuracy, a management mandate is the only way to institute widespread application of measurement assurance.

III.2 A GENERALIZED SELF-EVALUATIVE QUALITY SYSTEM

This section of the paper briefly describes the major steps in establishing a self-evaluative quality system and the generic elements of such a system which comprise the "criteria" for self evaluation. The quality system described here is based on Central Laboratory's experience with MAEs and NBS's experience with laboratory evaluation through NVLAP.

III.3 ESTABLISHING A SYSTEM

Determination of Need

Determination of the need for a MAE based self-evaluative quality system requires assessment of two major factors: the laboratory's measurement capability and the client's and/or laboratory management's needs for data of specific quality levels.

Complete and accurate definition of a laboratory's measurement capability is a prerequisite to establishing a self-evaluative laboratory quality system. The quality system may contain laboratory policies that do not relate directly to specific measurement capabilities, but the primary objective of the quality system must always be the sound and convincing evaluation of actual measurement quality. There is not an agreed upon way to present a laboratory's range of expertise and competence. Some laboratories categorize testing competence in terms relevant to the types of measurement that their client seek; some use generalized fields of testing. Still other laboratories categorize their competence according to personnel skills and the types of equipment available.

If a laboratory's clients demand precision and accuracy information, an economic driving force already exists to support MAEs. Clients need to consider the costs of poor accuracy as well as the benefits of high accuracy since measurement error affects profit, safety, and other costs. In engineering applications, cost savings may be identified by comparing the measurement variation to the specification limits. MAEs are less costly to implement with multiple analysts, multiple instruments and a high volume of similar measurements because the quality evaluation costs are then less per person, per instrument, and per measurement respectively.

Another type of economic driving force may be government regulations such as mandated air or water analyses where the risk of noncompliance carries significant economic penalties. Lacking these economic driving forces, laboratory management may decide to implement and support MAEs with laboratory policies and objectives. Central Laboratory's experience is that management's desire for measurement assurance is insufficient to support and maintain MAEs without client demand. As a long-term strategy, laboratory management should commit funds, motivational support, and leadership in the statistical education of clients to build future client support for a quality system. In the near term, any management decision to implement MAEs must balance immediate costs against potential benefits and the risks associated with data of unacceptable quality. Congruence with long-term laboratory goals and plans must also be considered.

Selection of Measurements

After deciding to establish a quality system, a list of measurement requiring MAEs must be compiled. This list provides the basis to evaluate the magnitude of the task. For laboratories that conduct a small number of different measurement, the task will be quite different from that for a broad-based laboratory conducting a large number of different measurements with varied frequency. The factors to be considered in selecting measurements are: ongoing sample volume, client demand for precision and accuracy information, the economic importance of the measurement, future measurement needs, available manpower to conduct applicable MAEs, and the cost of implementation. Measurements whose errors are associated with large risks will be given higher priority in this decision-making process. Regardless of the laboratory, the list of measurements provides the framework for measurement assurance. Cost estimates and time bounded objectives for each MAE may also be required.

Establish Working Groups

The responsibility for implementing the MAE should be assigned to the analysts and management personnel who normally provide the measurements to laboratory clients. The work group, which is comprised of one or more analysts and a group leader, must be given the authority, resources, and opportunity to actively participate in all phases of the assignment. This group will establish measurement process control, evaluate capability, decide whether sufficient resources are available, and implement ongoing precision measurements. The work group should start with an analysis of its needs, including procedures, sampling, equipment and facilities, statistical methods, standards, manpower, timing, and participation in decision-making. Specific needs will vary with each analysis and the successful completion of the assignment - establishment of a MAE - will require the satisfaction of those needs by the work group itself. Training, human resources, or equipment needs may require the support and assistance of laboratory management.

Develop Measurement Assurance Plan

After satisfying its needs, the work group should develop a measurement assurance plan. The plan should focus on the procedure for conducting the measurement, standards for reference analyses, frequency of reference analyses, control charting procedures, and any potential sources of measurement error specific to the measurement. The plan for each of the four MAEs evolved after a set of initial measurements. It is important to obtain measurements based on statistical considerations to permit the decisions that follow. Once an appropriate number of measurements have been obtained, either

as a group or over time, the standard deviation can be computed in order to give a measure of the variability of the measurement process. Control charts should be established appropriate to the frequency or the number of replicates. If control of the measurement process is judged unsatisfactory, return to the plan or the needs analysis. If the process control is sufficient, an evaluation of precision can be made. A good plan can minimize false starts and make for a more productive experiment.

Implement MAE

Implementation of an MAE should follow an experimental plan which defines the equipment and facilities to be used, sampling and material preparation procedures, and measurement procedures including calibration, data analysis, and reporting requirements (with significant findings and recommendations). The four MAEs described in the appendix are models of the technical and statistical approach to performing a MAE.

III.4 GENERIC QUALITY SYSTEM ELEMENTS (CRITERIA)

Laboratory management can choose to evaluate its competence to perform specific measurement and the "fitness for use" of its test data by:

- . assigning the responsibility for determining quality to an outside third party (e.g., accreditation or certification agency), or
- . self-evaluating quality by comparison to internally established criteria

After identifying those individual measurements which need an explicit demonstration of measurement capability by means of MAEs, laboratory management should consider those factors necessary to successfully implement the MAEs. The systematic application of formal self-evaluation should be based on the following general criteria (described briefly below and in more detail in appendix B.2) and will provide a mechanism to review and evaluate what exists and to plan for improvements if necessary.

Organization

Organizational criteria include development of an internal company policy for laboratory quality. A stated quality policy is essential for both staff and clients to recognize that management is committed to laboratory quality. Laboratory management must provide sufficient resources to implement and maintain the policy. Any such laboratory policy will be a function of organizational climate, management style, business practices, ethical policies, and both short and long term objectives. The quality policy provides direction for the implementation of quality assurance which is ultimately a management need.

The next three general criteria (personnel, facilities, and procedures) directly influence the quality of laboratory measurements. These elements interact with the quality control procedures as illustrated in exhibit III.

Personnel

Personnel criteria include the technical skills of laboratory staff, training procedures, professional interactions, responsibilities, accountability, and supervision. Depending on the nature of the testing conducted, these

personnel criteria may vary with measurement complexity. The degree of technical skill and/or training required will increase in proportion to the complexity of the test conducted. A self-evaluative laboratory quality system must include procedures to review staff performance, provide feedback, suggest training, and encourage professional interactions.

Facilities

The laboratory facilities have a direct impact on the quality of the measurements produced. Suitable and properly maintained equipment or instruments are key to most laboratories. The necessity of environmental controls and calibration standards will depend on the specific type of testing conducted, but they can be critical factors in the analysis. Self-evaluative procedures are necessary to determine if the facilities used to conduct laboratory measurement are in proper order. These procedures require documentation of calibration procedures and maintenance. Over and above these equipment, environmental, and calibration factors is the basic need to meet suitable housekeeping and employee safety requirements which, while not directly related to measurement quality, have an indirect impact on laboratory employees and the laboratory image.

Procedures

Quality assurance procedures should include proficiency testing, round-robin testing, or interlaboratory test programs. These testing programs provide the opportunity to critically compare laboratory procedures and measurements with peers in other laboratories.

Other procedures need to be explicitly set out for sample records, data records, calibrations, test methods, specifications and/or standards, good laboratory practices, MAEs and control charts, if applicable. These are all important for the proper conduct of measurements. A self-evaluative quality system must provide for periodic review of these factors to assure that corrective actions are implemented where appropriate.

Reporting

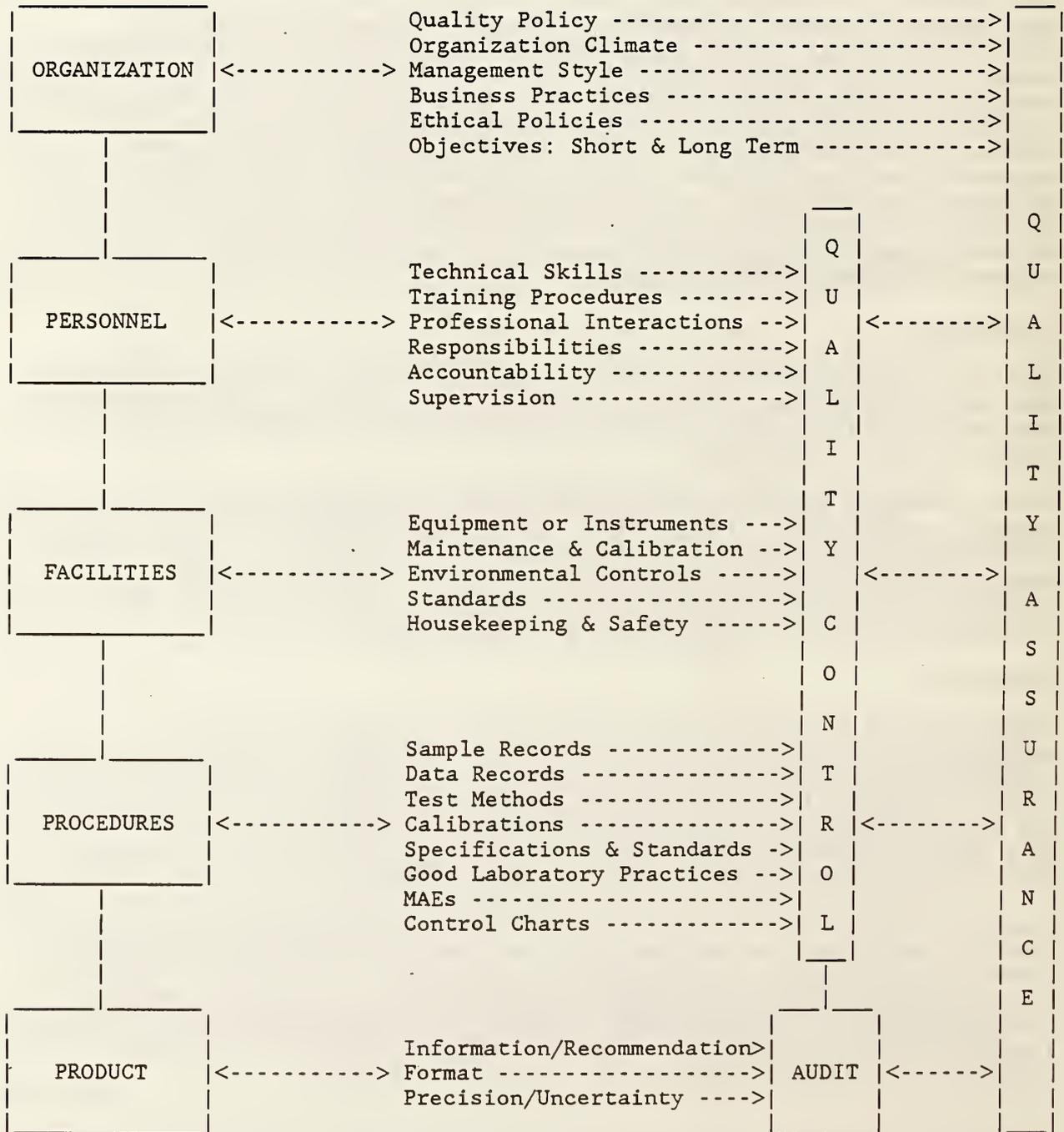
Some important features of test reports are the information and/or recommendations, format, and the precision of the reported measurements. The proposed self-evaluative laboratory quality system should rely on the internal audit of laboratory reports to identify quality concerns, make recommendations for procedural changes and interact with the operating quality control system. This dynamic internal audit provides the mechanism to close the loop between the quality assurance and the quality control subsystems.

Personnel, facilities and procedures all play a role in measurement quality. Quality control must be an interactive and dynamic process and it must be focused on the analysts conducting the measurements. Feedback to initiate corrective actions cannot wait until problems are detected by the quality assurance system. Laboratory staff conducting the measurement are the first link in a self-evaluative quality system and should have the necessary resources to initiate changes to the measurement process.

Exhibit III. - General Laboratory Quality System

GENERAL CRITERIA

Specific Activities



IV. REFERENCES

- [1] John K. Taylor, "Quality Assurance of Chemical Measurements," Seminar Materials, NBS, Washington, DC, 1983.
- [2] J. M. Juran, Ed., Frank M. Gryna, Jr., and R. S. Bingham, Jr., Quality Control Handbook, Third Ed., McGraw-Hill Co., NY, 1979, pp 2-24.
- [3] John Simpson, "Foundations of Metrology," Journal of Research, NBS, Vol. 86, No. 3, 1981, pp. 281-292.
- [4] "Quality Systems Terminology," ANSI/ASQC Standard A3-1978, 230 West Wells Street, Milwaukee, Wisconsin 53203.
- [5] ASTM E-548 Committee E36, Compilation of ASTM Standard Definitions, 5th Edition, ASTM, 1916 Race St., Philadelphia, PA 19103, p. 565.
- [6] George T. Moffatt, "New Initiative in Quality Assurance," AT&T Bell Laboratories Record, December, 1983, pp. 6-8.
- [7] James P. Dux, "Quality Assurance in the Analytical Laboratory," American Laboratory, July, 1983, pp. 54-64.
- [8] ASTM E-699 Committee E6, Compilation of ASTM Standard Definitions, 5th Edition, ASTM, 1916 Race St., Philadelphia, PA 19103, p. 564.
- [9] S. Gaft and F. D. Richards, "Quality Assurance at Ford Motor Company Central Laboratory: A Dynamic Approach to Laboratory Quality," ASTM E36 Symposium on Evaluation and Accreditation of Inspection and Test Activities, Washington, DC., April, 1981.
- [10] "Quality System Standard," Q-101, Product Quality Office, Ford Motor Company, 1983 Edition.
- [11] "Operator's Manual 1090 Thermal Analyzer", E.I. DuPont de Nemours and Company, Part Number 994280-001, May, 1982.
- [12] "Standard Practice for Compositional Analysis by Thermogravimetry", ASTM E-37 Committee Documentation, TM-01-09A-7, March, 1983.
- [13] "TGA Data Analysis, Compositional Analysis Version 2.0", E.I. DuPont de Nemours and Company, Part Number 994625-906, 1981.
- [14] Brian C. Belanger, "The Role of NBS Calibrations in Quality Assurance," ASQC Quality Congress Transactions, 1983, Boston.
- [15] W. W. Scherckenbach, "How to Train Employees in Statistical Techniques," ASQC 24th Annual Quality Clinic, Knoxville, TN, March, 1983.
- [16] Carroll Croarkin, Measurement Assurance Programs, Part II: Development and Implementation, NBS Special Publication 676-II, April 1984.
- [17] Carroll Croarkin, Statistical Engineering Division, NBS private communication, June 12, 1984.
- [18] "Instruction Manual 200-172, CS-144, Carbon & Sulfur 780-700 System," LECO Corporation, 3000 Lakeview Ave., St. Joseph, MI, 49085, 1979.

- [19] Dr. Kaoru Ishikawa, Guide to Quality Control, Nordica International Limited, Tokyo, Japan, 1982, pp 18-28.
- [20] Alpha Resources Inc., Box 199, 3090 Johnson Rd., Stevenville, MI, 49127
- [21] "Continuing Process Control and Process Capability Improvement," Statistical Methods Office, Ford Motor Company, July, 1983, p 14.
- [22] John C. Root, "Petro-Chem Technical Letter: Antifreeze-Coolants - A Review," Witco Chemical Corporation, January, 1983.
- [23] H. M. McNair, "Equipment for HPLC--V.", Journal of Chromatographic Science, Vol. 20, December, 1982.
- [24] Diago P. Lopes and James A. Galloway, "Analytical Note: An Ion-pair Chromato-graphic Technique for Analysis of Organic Additives in Engine Coolant," Journal of Chromatographic Science, Vol. 20, November, 1982.
- [25] Kenneth A. Brownlee, Statistical Theory and Methodology in Science and Engineering, J. Wiley & Sons, 1960, pp 255-269.
- [26] Carroll Croarkin, John Beers and Clyde Tucker, "Measurement Assurance for Gage Blocks," NBS Monograph 163, February, 1979.
- [27] J. M. Cameron, "Measurement Assurance," NBSIR 77-1240, April 1977.
- [28] John Simpson, "NBS Approach to Quality," Test and Measurement World, December, 1983, p. 38.

A. MEASUREMENT ASSURANCE EXPERIMENTS

A.1 THERMOGRAVIMETRIC ANALYSIS

This measurement assurance experiment (MAE) was intended to empirically determine the attainable precision and any bias between two Dupont 951 Thermogravimetric Analysis (TGA) systems. TGA data are used to fingerprint elastomers, plastics and paints and to establish master thermograms on approved materials. TGA data may also provide useful information for failure analysis of polymeric parts and materials. In spite of the heavy demand for TGA analyses, the precision and accuracy of this type of analysis was not previously documented. The two objectives of the MAE were: (1) to determine if the measurement process was in control and (2) to determine the attainable precision.

Technique and Equipment

TGA is a destructive analysis in which the weight change of several milligrams of a sample is measured as the sample decomposes during heating at a controlled rate. Typically, the sample is heated first in nitrogen to determine the weight loss of volatile components, then in air or oxygen to determine the weight loss of oxidizable components. For elastomers, typical components that volatilize in nitrogen are: (1) plasticizers, (2) polymers and (3) components from thermally decomposable fillers (e.g., CO₂ from CaCO₃). Typical components that oxidize in air are: (1) carbon black, (2) graphite and (3) carbonaceous residues from highly cross-linked polymers. A typical weight loss curve is shown in figure 1.*

The Dupont 951 Thermogravimetric Analyzer Balance Assembly is shown in figure 2. The specimen rests in a platinum sample pan, (A), which is suspended on the balance arm, (B), inside a quartz furnace tube, (C). A thermocouple, (D), near the sample pan, monitors the temperature. The atmosphere inside the quartz furnace is supplied through the purge gas inlet, (E), and is regulated by a valve. The entire quartz furnace tube slides into the furnace. The temperature, heating rate, specimen weight and gas flow rate are measured and controlled by the Dupont 1090 Thermal Analyzer, which consists of a computerized programmer, a dual disc drive, a plotter/printer and keyboard/plasma screen. The operation of the TGA system is described in the manufacturer's operating manual [11].

Experimental Plan

The analytical variables include the sampling, the procedure, the analysts, the sample, and the equipment. To "monitor" all variables, the sampling had to be representative of regular analyses. The method used follows ASTM E-37 for thermogravimetric compositional analysis [12], modified as follows: a sample size between 18 and 22 milligrams, a cool down step was added to the heating program, and all calculations are performed by data analysis program [13]. Four analysts conducted the planning, implementation and data collection. All decisions concerning the test procedure, type of sample, sample size, test frequency and data retention were made by consensus. The parameters recorded for each analysis for possible later identification of sources of error were: sample size, furnace used, visual condition of the platinum, history of platinum pan use, visual condition of the thermocouple,

*Figures are shown beginning on page 42

gas flow rate, any maintenance or repairs, and all calibrations. Each analyst performed one 90 minute analysis per week on alternate TGA systems. Control charts on both TGA systems were selected to obtain a precision statement and to identify any bias between the two systems. The control charts were also used to monitor the ongoing measurement process.

An elastomer sample was arbitrarily chosen as a check standard for this project because elastomeric standard reference materials or known secondary reference standards are unavailable. An acrylonitrile elastomer, Polysar Limited Automatic Transmission Reference Rubber ATRR 100-P1 - batch 3, was chosen because it is similar to samples routinely analyzed; it is used as a reference material for testing automatic transmission fluids and as a round-robin sample for tensile tests. Homogeneity of the check standard was a concern because it becomes part of the overall precision statement. To assure compositional stability of the check standard, the specimens were stored away from all light sources. Compositional variability of materials normally analyzed is expected to be worse than for the check standard.

The specimens for analysis were die-cut from a sample slab of the check standard with an ASTM D-746 modified T-50 die and trimmed to the desired 18-22 mg weight with a scalpel. The 18-22 mg specimen was placed on the platinum sample pan, sealed in the quartz tube and inserted in the furnace. The specimen was heated at a rate of 20°C per minute from 30°C to 900°C and then cooled to 600°C in a nitrogen atmosphere. The atmosphere was changed to air and the specimen reheated from 600°C to 900°C at 10°C per minute. Weight, time and temperature were automatically collected during the test and stored on disc for later analysis. Each analysis produced a thermogram (figure 1) from which three compositional parameters (polymer, carbon black and ash) were calculated by the computer and entered on the control chart.

Results

Data accumulation for the check standard began in May, 1983 and continued through November, 1983. Run charts based on consecutive runs for each system were set up to monitor the progress of the MAE. The data points collected are shown in exhibit A.1.1. Special conditions, such as major repairs or changing the platinum pan, were recorded. The mean and standard deviation for percent elastomer including plasticizer, percent carbon black and percent ash (residue) were calculated for each system after twenty-six runs. The upper and lower limits were calculated at three times the standard deviations and are shown in exhibit A.1.2. Lines corresponding to the three mean values and upper and lower limits were then added to the run charts. Data accumulation continued until the end of 1983 to see if any of the data points exceeded these upper and lower control limits. The control limits were exceeded on the thirty-first analysis using system one. A weight gain of approximately 3%, of unknown origin was observed at fifteen minutes. The control chart readily indicated the faulty data. The cause of the out-of-control condition (run 31, figure 3), however, could not be determined. The one instrumental parameter that affected the compositional data was the condition of the platinum pan. For this reason, the sample pan should be inspected before each analysis for signs of any deterioration and cleaned over a burner to remove any residue.

Thirty-six analyses were completed on system 1 by November, 1983, and 37 analyses were completed on system 2 during the same time. The out-of-control analysis was excluded from the statistics. The calculated means and three-sigma values for all analyses are shown in exhibit A.1.3. These standard deviations provide a measure of the analytical precision. The

means shown in exhibit A.1.3 illustrate a slight bias between the two systems. The average elastomer content obtained with system 2 is 0.38% larger than for system 1. (figure 3) The average ash content obtained with system 2 is 0.64% lower than for system 1. (figure 4) The carbon black percent is about the same for both systems; 45.45% and 45.64% respectively. (figure 3) The variability of the ash content, as measured by the standard deviation, is larger for system 2 than for system 1. For elastomer and carbon black, the standard deviations differ slightly.

Graphs were plotted and regression coefficients were computed for each component versus sample weight and each pair of components (e.g., carbon black versus elastomer and ash versus elastomer). The regression coefficients are shown in exhibit A.1.4. The analysis for ash versus sample weight for system 1 showed a downward trend that was confirmed by a t-test value of -2.56 for the linear regression coefficient. A graph of this regression analysis is shown in figure 5. No explanation for this correlation between ash and sample weight has been identified. The relationships between components indicates that both the % carbon black and the % ash decrease as the % elastomer increases and was verified by a t-test of linear regression coefficients. These relationships are illustrated in figure 6 for % carbon black versus % elastomer and figure 7 for % ash versus % elastomer. These correlations are expected because the sum of the components must equal 100%.

Plans for Future Action

Control charts show that the TGA measurement process is in control. All of the analysts are confident that the demonstrated variability is satisfactory for our application. With this information, the decision was made to modify the control procedure. Because the control chart is a tool to monitor the measurement process, a control analysis should be conducted every one to two months on each system to assure that the process remains in control. Since this MAE was deemed a success and can be easily monitored, extension to another related technique, differential scanning calorimetry, for determining the glass transition of elastomers or melt points of plastics is being considered.

Collaborative round-robin testing would be the next step in measurement assurance. For analyses such as TGA, without standard reference materials, this is the only way to establish an estimate of a generally accepted true value based on a comparison with analyses developed by consensus. Because control of the TGA measurement process has been judged sufficient, round-robin testing would permit evaluation of the ability of our TGA systems to produce results in agreement with other qualified laboratories.

Conclusions

The data produced by these TGA systems has sufficient precision for our use and the TGA measurement process is under control. These two factors give the analysts confidence in the measurement process and in the analysis technique. The analysts know that measurements for laboratory clients are reliable.

Two other factors were observed: (1) participative decisions were important to the experiment and (2) the MAE was useful in learning the analysis procedure. These benefits should not be undervalued.

Exhibit A.1.1 - TGA Polymer Analyses

System 1

System 2

Run*)	mg	% Elast.	% C	% Ash	mg	% Elast.	% C	% Ash
1	18.81	47.71	45.57	6.65	20.45	48.75	46.03	5.26
2	21.52	48.48	45.32	6.37	19.82	48.81	45.98	5.29
3	20.33	48.47	45.51	6.15	19.98	49.22	45.84	4.96
4	20.09	48.26	45.38	6.38	21.09	48.74	46.03	5.38
5	21.65	48.42	45.69	6.10	20.24	48.96	45.91	5.25
6	17.15	48.93	45.01	6.27	20.33	48.96	46.14	4.95
7	18.15	48.22	45.07	6.83	19.65	49.38	45.77	4.92
8	19.03	49.03	45.07	6.07	19.36	48.84	45.72	5.42
9	19.88	48.86	45.27	5.89	17.07	49.58	46.29	4.18
10	18.67	48.52	45.64	5.93	19.83	49.32	45.75	4.95
11	18.46	48.70	45.33	6.01	18.59	48.50	45.51	6.02
12	17.95	48.72	45.45	6.00	19.02	49.11	45.34	5.70
13	22.25	48.93	45.42	5.80	18.88	49.00	45.86	5.20
14	19.97	48.77	45.41	5.95	20.14	48.62	45.84	5.58
15	21.98	48.81	45.54	5.69	19.20	49.18	45.50	5.37
16	18.59	47.48	46.47	6.17	19.98	49.24	45.36	5.56
17	18.21	48.70	45.12	6.38	21.83	49.34	45.60	5.15
18	19.39	49.04	44.82	6.25	21.43	48.66	46.23	5.19
19	18.77	49.23	44.85	6.08	21.50	49.08	45.58	5.38
20	21.55	49.37	45.34	5.37	18.90	49.75	45.18	5.06
21	19.14	49.00	44.60	6.43	18.86	49.78	45.10	5.19
22	20.21	48.96	44.96	6.14	20.92	48.99	45.29	5.76
23	18.64	48.15	46.32	5.73	21.12	49.40	45.32	5.35
24	18.88	48.90	45.63	5.60	18.85	49.61	44.76	5.71
25	19.63	48.52	45.63	6.02	19.01	49.06	45.37	5.56
26	18.84	48.52	45.68	5.86	20.86	48.54	44.92	6.57
27	21.16	48.65	45.82	5.54	19.93	49.09	45.63	5.31
28	20.41	49.27	45.61	5.27	20.82	48.22	45.62	6.26
29	19.94	48.91	46.04	5.20	21.76	49.28	44.93	5.78
30	18.25	48.07	45.83	6.45	21.29	49.62	45.84	4.49
31	**18.75	**45.16	**46.79	**8.19	19.39	48.81	45.72	5.47
32	21.06	48.69	45.66	5.70	19.24	49.11	45.71	5.17
33	20.43	48.85	45.11	5.97	21.47	48.91	45.86	5.16
34	19.89	48.72	45.64	5.74	19.95	48.10	45.84	6.09
35	20.75	48.44	45.28	6.30	20.74	49.18	45.94	4.88
36	19.19	48.12	45.51	6.51	18.17	48.26	45.64	6.23
37	---	---	---	---	21.13	48.86	45.81	5.38
Mean	19.68	48.64	45.44	6.02	20.02	49.02	45.64	5.38
Std.Dev.		0.409	0.390	0.372		0.403	0.358	0.468

*Chronological run number. Run number on system 1 has no correspondence with run numbers on system 2.

**Out-of-control data was not used in standard deviation calculation.

Exhibit A.1.2 - Average Composition from 26 Runs

	System 1	System 2
Percent Elastomer		
Mean	48.64	49.09
3 x Standard Deviation	1.30	1.08
Percent Carbon Black		
Mean	45.39	45.62
3 x Standard Deviation	1.25	1.20
Percent Ash		
Mean	6.08	5.34
3 x Standard Deviation	0.96	1.31
Runs	26	26

Exhibit A.1.3 - Average Composition from 36 Runs

	System 1	System 2	Difference
Percent Elastomer			
Mean	48.64	49.02	0.38
3 x Standard Deviation	1.23	1.21	
Percent Carbon Black			
Mean	45.44	45.64	0.20
3 x Standard Deviation	1.17	1.07	
Percent Ash			
Mean	6.02	5.38	0.64
3 x Standard Deviation	1.11	1.40	
Runs	35	37	

Exhibit A.1.4 - Regression Coefficients for Elastomer Components

	System 1	System 2
Elastomer versus Sample Weight	0.082	-0.046
Carbon Black versus Sample Weight	0.020	0.012
Ash versus Sample Weight	-0.121*	0.029
Carbon Black versus Elastomer	-0.528*	-0.251*
Ash versus Elastomer	-0.476*	-0.770*

*Significant Regression Coefficients

A.2 GAGE BLOCK TOLERANCE TESTING

The objectives of this measurement assurance experiment (MAE) for gage block tolerance testing are to determine the precision attainable, evaluate if the precision is "good enough" and develop a plan for statistical control of the measurement process. Because the uncertainties associated with the calibration of gage block masters must be included in the total measurement uncertainty, this report also details the procedure to transfer NBS calibrations to working masters, and to determine the uncertainty associated with this transfer.

Gage blocks are small rectangular metal blocks with a well defined thickness used to calibrate micrometers, calipers, and other dimension measuring tools. Gage blocks are used to check or inspect manufactured components to determine conformance to dimensional specifications. A typical set of gage blocks consists of 81 blocks from 0.05" to 4" thick. Less often used gage blocks may be as large as 20".

The requirements for uncertainty and traceability of gage block inspections or calibrations depend on service requirements. Manufacturing concerns using gage blocks for critical dimensional measurements associated with NASA, Atomic Energy, or military applications may require small uncertainties with traceability to calibrations provided by NBS [14]. The uncertainty and traceability requirements for other manufacturing applications are usually less stringent. Central Laboratory provides tolerance testing of gage blocks for Ford Motor Company operations. Tolerance testing is not a calibration procedure, but each block is compared to blocks from a higher level reference set to determine if the given block's dimension is within a tolerance range of the nominal size. The tolerance requirements at Ford Motor Company are that blocks be within 4 millionths of nominal size for "A" quality or within 8 millionths of nominal size for "B" quality. If the block's dimension is within the tolerance range, it is returned to service; if not, the block is replaced or repaired.

Experimental Plan

When tolerance tested, the dimension of each gage block in an 81-piece set is compared to the equivalent sized working master using a comparator. The working master is placed in the comparator and the reading adjusted to zero or the calibrated value. The gage block to be tested is placed in the comparator and the difference determined. This process gives a difference from the nominal size. To initiate control and estimate the current precision for this tolerance testing process, ten gage blocks typical of the range of sizes measured were selected from inventory for use as measurement controls. Because both metric and English gage blocks are tested, control blocks representing each were necessary. The dimensional difference between the working master and these ten control blocks, in millionths of an inch or micrometers, was recorded each time a production gage block of that size was tested. Because the tolerance testing is conducted by two inspectors using two different comparators, the date, time, inspector, temperature and humidity were also recorded.

Results

Typical data from this process for the English control blocks, one inspector, one comparator and one working master set are shown in exhibit A.2.1. The

measurements on the 1.0" English control block showed that the average difference between the 1.0" working master and the 1.0" control block is 6.4 microinches (0.0000064") with a standard deviation of 3.0 microinches. The precision of a single measurement, estimated as three times the standard deviation, is ± 9.0 microinches for testing the 1.0" block. This range applies to the tolerance testing process with one inspector, one comparator and one English working master set. Similar control block data for the second English working master set and a different inspector and comparator are shown in exhibit A.2.2.

Using the measure of precision developed from the process control measurements, the next task is to evaluate if the measurements are "good enough." For an "A" quality 4.0" block, the tolerance range is ± 0.000016 " or a total range of 0.000032". The measured precision of ± 0.0000339 " for working master set 1 from exhibit A.2.1, is significantly larger than the entire tolerance range, therefore, the precision is insufficient. Even the tolerance requirement for "B" quality of ± 0.000032 " is not satisfied with a precision of ± 0.0000339 ". This evaluation of precision for the control blocks compared to master set 1 is shown in figure 8 and compared to master set 2 in figure 9. For the working master set 2, "A" quality tolerance is feasible for blocks smaller than 4.0". For set 1 (figure 8), even "B" quality tolerance is not feasible for the entire range of sizes. Notable exceptions are the 4.0, 2.0, 1.0 and 0.125 inch blocks. The effect of temperature fluctuations may be a significant factor for the larger gage blocks. Refinement of the gage block tolerance testing process will be necessary to improve this precision.

Similar control block data from process controls for metric gage block tolerance testing with one inspector, one comparator and one metric working standard are shown in exhibit A.2.3 and figure 10. Excellent precision for the entire range of metric block sizes is demonstrated satisfying both "A" and "B" tolerance requirements. In summary, the tolerance testing process for metric gage block tolerance testing is in control and is producing measurements that are "good enough."

Plans for Future Action

A basic principle of process control is that the precision of a process in control cannot be significantly altered except with major changes in the process [15]. This raises the question of what should be changed. To efficiently determine the most significant factors influencing the variability of the measurement process, a statistically designed factorial experiment is proposed. The three factors would be the inspectors, the comparators, and the working masters. Each of these factors have two possible levels. Thus, an experiment with three factors, each at one of two levels would be conducted. To provide a degree of confidence in the total experiment, each measurement would be repeated four times. After the significant sources of variability have been identified, changes to the measurement process can be contemplated (e.g., equipment, personnel or environmental controls). Then the cost of better precision can be evaluated. This factorial experiment and changes to the measurement process will not be covered in this paper, but will influence future decisions on gage block tolerance testing.

Calibration of Working Masters

The measure of total uncertainty for the tolerance testing process must include the uncertainty in the working master calibration and the uncertainty in NBS calibrations. The second part of the measurement assurance experiment evaluates the process used to calibrate the working masters used for the tolerance testing. This calibration process was set up following the procedure described in Croarkin [16]. Briefly, the process includes an inter-comparison between two sets of NBS calibrated gage blocks and the working masters.

The measurement process for the 81 gage blocks in a set is as follows:

- . the working master, X, is placed in the comparator and the comparator is set to zero. The X1 reading is zero.
- . NBS master S1, is measured and recorded.
- . NBS master S2, is measured and recorded.
- . the working master, X, is again measured and X2 is recorded.
- . the entire process is completed twice for each working master by each inspector, a total of four complete measurements.

A sample of measurements obtained with this procedure is shown in exhibit A.2.4. The control of this measurement process is maintained by monitoring the measured difference between the two NBS calibrated masters, S1 and S2 as described in [16]. Another check on the measurement process control is shown in exhibit A.2.5, where the standard deviation for a group of four separate measurements of the difference between S1 and S2 are compared to the pooled standard deviation for groups of measurements on gage blocks of comparable size.

The F-test shown in exhibit A.2.5 was used to identify measurements that have excess variability. This process was especially helpful in identifying a problem with one 4.0" working master. Repeated measurements for the 4.0" working master continued to fail the F-test. It was noticed, however, that the difference between the size of the NBS masters and the working master approached the extreme range of the comparator reading. This raised the question of linearity at the extremes of the comparator readings. To eliminate this source of error, the 4.0" working master was replaced with another block, and the calibration process was then repeated. The calibration of the different 4.0" working master, closer in size to the NBS masters, was successful.

The total uncertainty in the working master calibration is computed by the following equation [17]:

$$U = 3s_p/2(n)^{1/2} + (U_{S1} + U_{S2})/2$$

with $n = 4$, s_p the pooled standard deviation of the difference between S1 and S2 as defined in [16] for groups of similar sized blocks, and U_{S1} and U_{S2} the uncertainties in the NBS calibrations. For example, for the 0.105000" block shown in exhibit A.2.7, this calculation gives a total uncertainty in the calibration of 4.7 microinches. This uncertainty in the calibration must be added to the precision of the tolerance testing process, which for a block of this size is 7.8 microinches from exhibit A.2.1 or 1.5 microinches from exhibit A.2.2. Further reduction of the uncertainty in the calibration of working masters is dependent on improving the precision of the entire measurement process.

Conclusions

The use of control measurements for the gage block tolerance testing process has provided a measure of precision of the process. This measure of precision makes possible valid decisions about whether the measurements are "good enough." The remaining task is to determine the cost of more precise measurements and whether additional cost is warranted.

The determination of the entire process uncertainty awaits the results of further study. Identification of the significant sources of measurement error should suggest improvements that can be made in the process. However, the procedure described in [16] to calibrate the gage block working masters has been implemented. Once the measurement process variability has been reduced, this process could be used for participating in an NBS Measurement Assurance Program (MAP), which may offset the cost of annual NBS calibrations.

Exhibit A.2.1 - Difference in Microinches Between
Control Blocks and English Working Master Set 1

Date\Inches	4.000	2.000	1.000	0.750	0.500	0.250	0.140	0.125	0.100	0.050
4/22/83	34.0	10.0	12.2	8.0	7.0	3.3	1.8	-3.8	7.8	-2.0
5/23/83	20.0	1.0	5.0	7.0	7.0	1.0	-1.0	-2.0	-1.0	-1.0
6/6/83	18.0	8.0	10.0	8.0	8.0	2.0	0.0	-1.0	-2.0	-2.0
6/6/83	4.0	1.0	6.0	5.0	8.0	2.0	1.0	5.0	-2.0	-2.5
7/6/83	-9.0	-12.0	6.0	4.0	6.0	0.0	1.5	-1.0	0.0	-2.0
9/22/83	1.0	9.0	6.1	6.5	6.0	3.0	3.5	7.0	3.0	-3.0
9/23/83	6.0	4.0	7.0	7.5	6.5	3.5	3.0	5.5	2.3	-3.2
9/26/83	8.0	4.0	6.0	6.0	6.0	2.5	4.4	6.0	0.7	-1.0
10/20/83	-2.1	8.0	6.0	7.0	5.0	2.0	3.0	7.0	1.0	-2.0
11/9/83	12.0	3.0	8.0	8.0	7.0	3.0	1.0	6.0	2.0	1.8
2/22/84	8.0	4.0	0.0	-1.0	4.5	1.0	0.0	5.0	1.0	-3.0
4/2/84	4.0	1.0	4.0	7.0	5.0	2.5	2.0	4.0	1.0	-1.0
AVERAGE	8.3	3.4	6.4	6.1	6.3	2.2	1.7	3.1	1.2	-1.7
Std.Dev.	11.3	5.8	3.0	2.5	1.1	1.1	1.6	3.9	2.6	1.4
3x Std.Dev.	33.9	17.4	9.0	7.5	3.3	3.3	4.8	11.7	7.8	4.2

Exhibit A.2.2 - Difference in Microinches Between
Control Blocks and English Working Master Set 2

Date\Inches	4.000	2.000	1.000	0.750	0.500	0.250	0.140	0.125	0.100	0.050
4/28/83	4.0	7.0	7.0	8.0	7.0	1.0	1.0	4.0	1.0	-1.0
7/12/83	14.4	11.0	9.0	10.0	9.0	1.0	1.7	3.9	1.2	-2.0
7/13/83	13.4	5.0	9.0	10.0	7.4	1.0	3.0	4.0	1.2	-1.6
7/29/83	13.2	6.4	11.1	7.4	8.3	3.4	4.3	4.4	1.0	-2.6
2/13/84	1.8	10.0	7.2	7.9	7.1	4.0	3.3	2.7	0.0	-0.3
AVERAGE	9.4	7.9	8.7	8.7	7.8	2.1	2.7	3.8	0.5	-1.5
Std.Dev.	6.0	2.5	1.7	1.2	0.9	1.5	1.3	0.6	0.5	0.9
3x Std.Dev.	18.0	7.5	5.1	3.6	2.7	4.5	3.9	1.8	1.5	2.7

Exhibit A.2.3 - Difference in Micrometers
Between Metric Control Blocks and Working Master

Date\mm	1.3	2.5	3.0	3.5	6.5	12.5	19.0	25.0	50.0	100.0
5/16/83	0.11	0.10	-0.05	0.06	0.00	0.05	0.06	-0.05	0.00	0.14
6/3/83	0.10	0.15	-0.01	0.14	0.04	0.06	0.00	0.02	0.14	0.20
6/23/83	0.11	0.13	-0.04	0.16	0.02	0.07	0.12	-0.02	0.01	0.15
7/29/83	0.12	0.12	0.02	0.17	0.00	0.10	0.09	0.03	-0.01	0.19
9/30/83	0.12	0.18	0.01	0.16	-0.01	0.08	0.11	0.00	0.12	0.20
1/10/84	0.12	0.13	0.01	0.15	0.02	0.07	0.13	-0.06	-0.01	0.17
1/19/84	0.12	0.13	0.00	0.15	0.01	0.07	0.11	0.04	0.16	0.31
3/1/84	0.17	0.09	0.03	0.18	-0.01	0.08	0.06	0.04	0.01	0.27
AVERAGE	0.12	0.13	0.00	0.15	0.01	0.07	0.09	0.00	0.05	0.20
Std.Dev.	0.02	0.03	0.03	0.04	0.02	0.01	0.04	0.04	0.07	0.06
3x Std.Dev.	0.06	0.09	0.09	0.12	0.06	0.03	0.12	0.12	0.21	0.18

Exhibit A.2.4 - Working Master Calibration

Nominal	X1	S1	S2	X2
0.100900	0.0	1.2	2.8	0.0
0.101000	0.0	4.0	4.4	0.8
0.102000	0.0	1.8	4.0	-0.2
0.103000	0.0	3.3	6.0	-0.2
0.104000	0.0	2.7	4.2	-0.1
0.105000	0.0	0.8	3.0	0.2

Exhibit A.2.5 - F-test of Working Master Calibration

Nominal	Ave. S1-S2	s_i	s_i^2/s_p^2
0.100900	0.65	1.07	0.38
0.101000	0.80	0.35	0.04
0.102000	-2.73	0.13	0.01
0.103000	-0.78	3.73	4.63 *
0.104000	-0.60	1.77	1.04
0.105000	0.07	2.07	1.42
0.106000	-5.93	4.05	5.48 *
0.107000	-0.17	1.03	0.35

*F-test is significant at 1% probability level if s_i^2/s_p^2 is greater than 4.2 with pooled standard deviation of 1.73 microinches.

A.3 LECO, CARBON ANALYSIS

The purpose of this measurement assurance experiment (MAE) is to empirically determine the precision and bias of the LECO carbon analysis of cast iron using a control chart. The precision was estimated from the upper and lower control limits for a repeatedly run control sample. The bias was estimated by comparing the empirically determined value for a Standard Reference Material (SRM) with its certified value.

Our laboratory analyzes hundreds of cast iron samples annually and carbon is a key element. Carbon in cast iron is an ideal candidate for a control chart experiment measuring precision because: (1) a large number of samples are analyzed and (2) carbon is key to material suitability decisions. These carbon analyses are used to make decisions about material suitability by comparing the measured carbon content with engineering specifications. The decisions might be made more easily if a good measure of precision were available.

There are several known sources of uncertainty in the chemical analysis of cast iron. In cast iron, carbon is present in two forms: graphite and carbide. The presence of carbon in these two forms creates an inherent uncertainty in sampling. The graphite can segregate when samples are prepared by drilling rather than cutting small pieces. Sampling from the heat affected surfaces instead of the core can also introduce uncertainty. Analyst training, procedures, instrument maintenance, and instrumental eccentricities are other sources of uncertainty in the measurement process. The accuracy of the analysis is also affected by any compositional uncertainty in the secondary or Standard Reference Materials (SRM) used for instrument calibration. The carbon analysis MAE demonstrates through a control chart whether all these sources of uncertainty are under control. This section explores these uncertainties, shows the calculation of precision from the control chart, and summarizes the benefits of a MAE.

Equipment and Operation

The equipment used for this MAE is a LECO CS-144 simultaneous direct reading carbon/sulfur determinator. This equipment has an automatic crucible loader, an integral electronic balance, and a digital printer. The LECO Corporation carbon analyzer can analyze small solid pieces of cast iron, high or low carbon steel, alloy steel, slag, or refractories for carbon content from 0 to 99.9%. The basic principle of the LECO carbon analysis is to burn a sample in a stream of purified oxygen, followed by the infrared determination of the carbon dioxide combustion product.

The analysis process begins by purifying pressurized oxygen. Next, the sample in a ceramic crucible is introduced into an induction furnace and burned in the purified oxygen stream. The gaseous combustion products (carbon dioxide, sulfur dioxide and water) pass through a moisture trap to remove water. The remaining gases pass through an infrared (IR) cell for sulfur dioxide analysis. The gas, containing some carbon monoxide, is fully oxidized by a catalyst to carbon dioxide and sulfur trioxide and passes through a cellulose trap to remove the sulfur trioxide. The carbon dioxide passes through a flow-thru IR cell. A heated nichrome wire emitting IR energy establishes a reference voltage in the 100% oxygen atmosphere and the carbon dioxide absorbs this IR energy. The output voltage of the flow-thru cell decreases proportionally to the amount of carbon dioxide present. A microprocessor reads the output voltage and compares it with stored values for that particular cell and detector. Stored factors (i.e., slope, intercept, noise

suppression, cell temperature, barometric pressure, flow rate, sample weight compensation, and blank and standard corrections) in the microprocessor are used to correct the data before the result is displayed as percent carbon. The complete operation of the LECO carbon/sulfur determinator is described in the manufacturer's operating manual [18].

Experimental Plan

The experimental plan for this MAE was developed with the voluntary participation of five analysts. The steps were to: specify the task, explore the experimental parameters, choose the most important analysis parameters, select the working standard, write a detailed procedure, and institute participative feedback to alter the plan as required. The specified task was to use a control chart to monitor the quality of LECO carbon analyses. The analysts chose the carbon concentration range for the control chart to be cast iron containing between 3.0 to 3.5% carbon.

The experimental parameters of the MAE were explored through cause and effect techniques [19] and with open discussions to permit timely modifications to the MAE. Concurrent with exploring the analysis parameters, the NBS research associate constructed the first average and range control charts for nineteen quadruplet analyses of NBS SRM 3c, containing 2.30% carbon. The control limits calculated from this data are shown in figures 11 and 12. Also, analyst training by the instrument manufacturer facilitated the exploration of the experimental parameters.

After the analysts were satisfied with the investigation of the experimental parameters, the calibration standard was chosen as the most important parameter. An NBS SRM standard was judged too expensive for an ongoing control sample, hence a secondary standard in the range of interest was selected. (Namely, Alpha Resources Inc. gray iron, AR302 [20], certified at $3.30 \pm 0.03\%$ carbon). An NBS SRM was run intermittently or when an out-of-control situation was noticed.

A procedure was written to detail the process of collecting data on the secondary standard for the control chart. The specific procedure was to run a set of triplicate analyses on the secondary standard and to plot the average for each point on the control chart. One set was plotted before and one set after each group of analyses. If the point was outside the control limits, the instrument was cleaned, drying agents changed, etc., and the instrument was re-standardized. For analyses of a large number of samples, this procedure was modified by plotting an additional secondary standard set in the middle of the analysis run. The decision on the number of samples analyzed between secondary standards was based on the number of samples the analysts were willing to rerun if an out-of-control situation occurred. Also, all values of the secondary standard, up to and including the value which fell within the control limits, were discarded during the morning warm-up period. In summary, a set of three analyses of a secondary standard were averaged for each point on the control chart. These control standards were run before and after a group of samples.

Results

This MAE was successful in measuring the precision and bias for the LECO carbon analysis. The control chart on the Alpha Resources Inc. AR302 secondary standard was implemented and the control limits for the average and the range were calculated when 20-25 sets of analyses had been obtained.

The most common method of calculating control limits for a control chart for averages (X chart) uses a measure of within-group variability (range or standard deviation). Exhibit A.3.1 illustrates an excerpt from a well-known statistical table used for this study [21]. Exhibit A.3.2 shows the calculation of three-sigma control limits based on 76 sets of LECO carbon data, using the formulas in exhibit A.3.1 and the average range of sets of three measurements. The three-sigma limit calculated from the within-group variability is $\pm 0.02\%$. The control charts for the average and range are shown in figures 13 and 14.

These limits, using within-group variability, were thought to be unrealistically narrow. Out-of-control points on the control chart occurred more than once out of every five points, instead of the three per thousand points expected for three-sigma limits. Repeated measurements, made without turning off the instrument, can give unrealistically narrow limits or unbelievably good precision. When significant and inevitable between-group variability exists, control limits for group averages should be calculated using the standard deviation of the group averages.

For the LECO carbon control chart data, which includes 153 individual measurements, the standard deviation is 0.028%. This estimate, multiplied by three is $\pm 0.084\%$, compared to the three-sigma limit of $\pm 0.026\%$ calculated from within-group variability. The large difference in precision calculated by the two methods is probably typical of measurement processes where repeated measurements are made without turning off the instrument. The larger sources of variation, setup and calibration are real, and must be included in a realistic estimate of precision.

The measured bias was less than the uncertainty of the NBS SRMs. Two additional discoveries were made in determining the precision and bias: a small instrument induced bias and an error in a certified value of an NBS SRM. The LECO instrument induced a bias of $+0.013\%$ carbon to the carbon results in the 3.30% range. The LECO instrument when calibrated with the 3.30% carbon secondary standard should give an average value on our control chart very close to the 3.30% certified value but instead averaged 3.313%. Figures 11 and 12 show the average and range control chart with limits calculated using exhibit A.3.2. An instrument calibrated with a specific standard is expected to give the calibrated value when analyzed as a sample. Although noteworthy, this bias was not judged significant when compared to the $\pm 0.05\%$ uncertainty of the NBS SRMs.

The plan to determine bias was to run a second less frequent control chart on a directly traceable NBS SRM and compare the average value with the certified value. NBS SRM 122g, with a value of 3.43% carbon, was chosen. The instrument was calibrated with the AR 302 secondary standard for this bias determination. The first empirically determined value of the NBS SRM 122g was 3.53% carbon: compared to the NBS given value of 3.43%; the bias is $+0.10\%$. Another SRM 122g, obtained from NBS, and SRMs 122d, 122f and 122h were analyzed. The carbon results from these SRMs were graphed on a calibration curve (figure 15), showed SRM 122g as an outlier. The footnote on the NBS SRM 122g certificate states expected deviation to be no more than $\pm 0.05\%$. NBS later confirmed that the first SRM 122g was in error. This MAE demonstrated the difficulty of determining any bias less than the uncertainty of the SRM, which for cast iron is $\pm 0.05\%$. However, the MAE provided sufficient confidence in the LECO carbon analyses to question the NBS certification of an SRM when a potential bias of $+0.10\%$ was observed.

Conclusions and Recommendations

In summary, the LECO carbon analysis MAE has an empirically determined three-sigma precision of $\pm 0.084\%$ and no detectable bias. This precision is larger than the manufacturer's statement of accuracy documented in the operations manual, " $\pm 0.5\%$ of the carbon present" (i.e., 3.30% carbon multiplied by ± 0.005 is $\pm 0.0165\%$). This difference is probably typical of manufacturer's claims which consider only the reproducibility of repeated measurements without turning off the instrument.

The direction which instituted the LECO carbon analysis MAE was removed to see if the analysts would continue control charting or if they would extend the practice to other analyses. Control charting stopped after approximately one year. The reasons cited were:

- . the control chart did not reduce normal maintenance or calibrations;
- . the reported precision was not appreciated by clients in spite of the efforts expended to explain the control chart information; and
- . the added time and cost inhibited using the control chart.

The MAE increased awareness of factors that may or may not affect measurement precision, a direct benefit of the MAE. For example, using the same exact weight of sample improved the precision. Conversely, using 99.6% pure oxygen did not improve the precision.

The final recommendation from this MAE is extracted from the above reasons for discontinuing the control charts. Laboratory management must make a commitment to support, recognize and, if necessary, mandate analysis precision until clients demand precision information.

Exhibit A.3.1 - Control Limit Calculations

The control limits are calculated to show the extent by which the subgroup averages and ranges would vary if only common causes of variation were present. They are based on the subgroup sample size and the amount of within-subgroup variability reflected in the ranges. Calculate the upper and lower control limits for ranges for averages:

$$\begin{aligned} UCL_R &= D_4R \\ LCL_R &= D_3R \\ UCL_X &= \bar{X} + A_2R \\ LCL_X &= \bar{X} - A_2R \end{aligned}$$

where D_4 , D_3 and A_2 are constants varying by sample size, with values for sample sizes from 2 to 10 as shown in the following partial table:

n	2	3	4	5	6	7	8	9	10
D_4	3.27	2.57	2.28	2.11	2.00	1.92	1.86	1.82	1.78
D_3	*	*	*	*	*	.08	.14	.18	.22
A_2	1.88	1.02	.73	.58	.48	.42	.37	.34	.31

*(for sample sizes below 7, the LCL_R would technically be a negative number; in those cases there is no lower control limit; this means that for a subgroup size 6, six "identical" measurements would not be unreasonable).

Exhibit A.3.2 - Upper & Lower Control Limit Calculations**

$$\begin{aligned} UCL_R &= D_4R = 2.57(0.025) = 0.064 \\ LCL_R &= D_3R = \text{Not Applicable} \\ UCL_X &= \bar{X} + A_2R = 3.313 + 1.02(0.025) = 3.339 \\ LCL_X &= \bar{X} - A_2R = 3.313 - 1.02(0.025) = 3.287 \end{aligned}$$

** Based on 76 data sets.

A.4 COOLANT INHIBITORS BY HIGH PERFORMANCE LIQUID CHROMATOGRAPHY

The analytical determination of tolyl triazole (TT) and 2-mercaptobenzothiazole (MBT) inhibitor concentrations using high performance liquid chromatography (HPLC) was selected for this measurement assurance experiment (MAE). The objectives were to determine the precision of the analytical method used to quantify these inhibitors in engine coolants and establish statistical control of the measurement process.

The HPLC data can fingerprint material composition for engineering specifications. Typical uses are to qualitatively characterize elastomers, vinyl plasticizers and polyols in urethane resins. A more frequent use of HPLC is the analysis of health-hazardous contaminants at the parts per million level in machining coolants. However, our most widely used application of HPLC data is the analysis of unused engine coolants containing a few grams per liter of TT or MBT inhibitor.

Engine coolant, commonly called antifreeze, is a mixture of ethylene glycol, propylene glycol, corrosion inhibitors, foam suppressant, dye and water added to radiators during final vehicle assembly [22]. Coolants from each assembly plant are analyzed quarterly for several chemical and physical properties to assure quality and proper in-process handling. Any deviations are resolved with the responsible supplier. Depending on the supplier, either TT or MBT is added to the coolant to inhibit corrosion of copper radiator components. The concentration of these inhibitors is measured by HPLC to assure that a minimum amount is present.

Technique and Equipment

HPLC is an analytical technique used to separate and subsequently quantify organic components in complex liquid mixtures. The separation occurs because each component flows at a different rate through a fractionating column. The flow rate is a function of the polarities of each component and its degree of adsorption on a fractionating column packed with a material of high surface area. These rate differences produce a different retention time for each component. The separated organic components are then detected with an ultraviolet photometer as they elute from the column. Each peak represents a component and the peak area is proportional to the amount present; the time axis shows the retention time. Quantitative results are calculated from integrated peak areas and calibration factors established from standards.

HPLC instrumentation, assembled as in figure 16, consists of the following components: solvent reservoir to contain the mobile liquid phase, high pressure pump to push the mobile solvent phase through the column, pressure gauge, automatic sampler, injector to introduce the samples into the mobile solvent phase, column packed with silica gel, detector to measure the concentration of each component, recorder and data handling module to record and calculate the concentration [23].

The equipment used for this MAE on engine coolants is a Waters Associates high performance liquid chromatograph equipped with: a WISP 710B automatic sampler, a U6K injector, a model 6000A solvent delivery system, a model 440 ultraviolet absorbance detector, and a model 730 data module.

Experimental Plan

The general method used for the analysis of TT and MBT in engine coolants is a published ion-pair chromatographic technique requiring minimal sample pre-treatment [24]. The HPLC is calibrated by injecting a laboratory prepared synthetic coolant made by mixing specific amounts of reagent grade sodium tetraborate ($\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$), sodium phosphate tribasic ($\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$), sodium benzoate, TT, MBT, ethylene glycol, distilled water, and phenol as an internal standard.

Samples for analysis are prepared by mixing 1 mL of coolant with 1 mL of a phenol solution containing 500 mg phenol/100 mL distilled water, and diluting to 10 mL with distilled water. Five microliters of the sample is injected into the HPLC mobile phase consisting of a mixture of 55% by volume tetrabutyl ammonium phosphate (0.00125M) and 45% acetonitrile, both HPLC grade. The coolant sample and mobile phase pass through a Bondopak C 18 chromatographic column, and the TT or MBT are separated from the other coolant components. The TT or MBT is detected with an ultraviolet photometer at 254 nm and the peak is recorded on a chromatogram. The corresponding concentration in grams of inhibitor per liter of coolant is calculated by the microprocessor. A chromatogram showing the location of TT and MBT peaks in a synthetic engine coolant is shown in figure 17.

Two coolants that contain typical concentrations of TT and MBT were selected as reference samples for this MAE. Union Carbide coolant was used for the TT reference and BASF Wyandotte coolant was used for the MBT reference because they contain typical mid-range amounts of the inhibitors.

For determining the precision of the HPLC measurement of TT and MBT, three precision studies were performed. In each study, eight repetitions of the basic measurement sequence were made on one reference sample and three quarterly coolant samples (32 samples - 8 aliquots of 4 samples). Sample preparation and instrument calibration with the laboratory prepared synthetic coolant were performed prior to each precision study. For each study, 32 samples were loaded in the autosampler and sample injection, separation, detection and calculations were all performed automatically. In each study, eight measurements were obtained for the reference sample and three quarterly coolants.

The TT reference standard was measured in the first two studies, and the MBT reference standard was measured in the third study. Two quarterly coolant samples containing TT and one containing MBT were measured in each study. The first two studies were conducted on consecutive days, and the third a month later.

Results

A total of 16 measurements were obtained for the TT reference sample and 8 measurements for the MBT reference sample. The quarterly samples, 1, 2, and 3, were analyzed in all three studies to obtain a larger amount of data for statistical evaluation. The quarterly coolant samples 1 and 2 contain TT so a total of 48 individual measurements of TT were obtained. Coolant sample 3, which contains MBT, was analyzed 24 times. The data obtained are shown in exhibits A.4.1, A.4.2, and A.4.3. The day-to-day and the within-day variability can be seen in figures 18, 19 and 20 for TT and figure 21 for MBT.

For the TT reference sample, the mean TT concentration is 0.59 g/L with a standard deviation of 0.011 in the first study, and 0.60 g/L with a standard deviation of 0.009 g/L in the second study. This variability for the TT reference sample, as measured by the standard deviation, was nearly the same for both studies. All individual measurements of TT were within ± 3 standard deviations.

The mean TT concentrations for coolant 1 are 1.15, 1.12, and 1.14 g/L in the three studies, respectively. The respective standard deviations are 0.009, 0.017, and 0.024 g/L. For coolant 2 the mean TT concentrations are 1.20, 1.20, and 1.16 g/L with standard deviations of 0.033, 0.012 and 0.020 g/L respectively. All individual TT measurements for coolant samples 1 and 2 were within the ± 3 standard deviations calculated for each study. However, the standard deviations for these coolant samples are slightly larger than the standard deviations for the TT reference sample.

To take advantage of the repetitive measurement of TT in the coolant samples, all of the standard deviations were pooled using the following equation to find the within-group component of variance:

$$\text{Pooled Std. Dev.} = \left[\frac{1}{k} \times \sum (S_i)^2 \right]^{1/2}$$

where $k = 8$, the number of groups pooled and S_i is the standard deviation of the individual groups. The pooled within-group standard deviation for TT concentration based on the 8 groups of eight measurements is 0.019 g/L. The standard deviation of the between-group component of variance was calculated using the following equation:

$$S_{\text{between pooled}} = \left[S_{\text{random pooled}}^2 - S_{\text{within pooled}}^2/n \right]^{1/2}$$

where $n = 8$, the number of measurements in each group. The $S_{\text{random pooled}}$ values were calculated using a One-Way Analysis of Variance Model [25]. Specifically, $S_{\text{random pooled}}$ was calculated by computing, for each sample, the standard deviation among the mean values for each precision study, and then pooling the results for all TT samples. The $S_{\text{between pooled}}$ for TT was 0.017 g/L. The estimated precision for a single analysis is obtained by combining $S_{\text{between pooled}}$ and $S_{\text{within pooled}}$ to form S_{total} as follows:

$$S_{\text{total}} = \left[(S_{\text{between pooled}})^2 + (S_{\text{within pooled}})^2 \right]^{1/2}$$

The value of S_{total} is 0.0255 g/L for TT. Using three times S_{total} as an uncertainty limit for TT measurements yields an uncertainty of ± 0.076 g/L which is adequate to establish the TT concentration for coolants containing 0.5 to 2.0 g/L.

For the MBT reference sample, the mean concentration obtained is 1.34 g/L with a standard deviation of ± 0.16 g/L. This standard deviation is an order of magnitude larger than the pooled standard deviation of ± 0.019 g/L obtained from the TT data. Because these analyses and inhibitor concentrations are similar, an equivalent standard deviation would be expected. Evaluation of the chromatograms identified interferences with the phenol internal standard peak in all eight MBT reference sample measurements. This unidentified component was not completely separated from the phenol peak as shown in figure 22. The amount of interference varied between the measurements and could not be identified without using other techniques. The internal standard interference prevented the accurate calculation of MBT values. Therefore, these measurements were not used to establish precision for the MBT analysis.

The mean MBT concentrations for coolant 3 are 2.94, 2.91 and 2.76 g/L in the three studies, and the corresponding standard deviations are 0.051, 0.060, and 0.036 g/L. All individual measurements of MBT in the quarterly coolants are within ± 3 standard deviations of the mean. In each measurement sequence the first MBT analysis had the lowest value. The significance of this observation, however, can only be established with further testing. The chromatograms show no interference with the internal standard or other coolant components that would make the data suspect.

The pooled standard deviation of the within-group component of variance for the MBT concentration based on the three groups of measurements obtained on coolant 3 is 0.050 g/L. The value of S_{between} based on the same 3 groups of measurements is 0.052 g/L. The calculated value of S_{total} is 0.072 g/L for MBT. Using three times S_{total} for the measurement of MBT gives a precision of ± 0.216 g/L which is adequate for typical MBT concentrations in coolants between 1.0 and 3.0 g/L.

Plans for Future Action

Data collection and ongoing control charts for the two reference samples will continue when quarterly coolants are analyzed to improve confidence in the precision measurement. The pattern of the first MBT analysis in each sequence being the lowest value will also be investigated. An experiment to estimate the accuracy will then be developed with interlaboratory testing. Documentation of the procedure will be performed to eliminate any analytical inconsistencies. The internal standard interference observed with the MBT reference coolant will be identified using Gas Chromatography/Mass Spectrometry.

Conclusions

The rigorous statistics used in this MAE provided an objective critical evaluation of the measurement process, and is being used as a tool to improve the measurements. For example, this MAE identified the need for precise documentation of the procedure, including evaluation of the chromatograms for peak interferences. Also, data should be plotted on control charts and evaluated by the analysts as generated to provide immediate feedback in problem identification. The delayed plotting and statistical evaluation in this experiment hindered analyst involvement and commitment to the experiment. The analyst training in the measurement process and in statistical procedures was a benefit that will be applied to future MAEs.

Exhibit A.4.1 - Precision Study 1

Measurement Sequence	Union Carbide Std. Reference TT, g/L	Quarterly Sample 1 TT, g/L	Quarterly Sample 2 TT, g/L	Quarterly Sample 3 MBT, g/L
1	0.58	1.14	1.19	2.88
2	0.59	1.14	1.22	2.96
3	0.59	1.15	1.20	3.02
4	0.60	1.15	1.23	3.00
5	0.60	1.15	1.14	2.91
6	0.58	1.14	1.19	2.95
7	0.61	1.16	1.25	2.91
8	0.60	1.13	1.21	2.89
Mean	0.59	1.15	1.20	2.94
Std. Dev.	0.011	0.009	0.033	0.051
3x Std. Dev.	0.033	0.027	0.099	0.15

Exhibit A.4.2 - Precision Study 2

Measurement Sequence	Union Carbide Std. Reference TT, g/L	Quarterly Sample 1 TT, g/L	Quarterly Sample 2 TT, g/L	Quarterly Sample 3 MBT, g/L
1	0.62	1.15	1.19	2.79
2	0.60	1.12	1.20	2.95
3	0.59	1.10	1.20	2.99
4	0.60	1.11	1.22	2.94
5	0.60	1.10	1.20	2.93
6	0.60	1.13	1.19	2.91
7	0.60	1.12	1.20	2.91
8	0.61	1.13	1.22	2.87
Mean	0.60	1.12	1.20	2.91
Std. Dev.	0.009	0.017	0.012	0.060
3x Std. Dev.	0.027	0.051	0.036	0.18

Exhibit A.4.3 - Precision Study 3

Measurement Sequence	BASF Wyandotte Std. Reference MBT, g/L	Quarterly Sample 1 TT, g/L	Quarterly Sample 2 TT, g/L	Quarterly Sample 3 MBT, g/L
1	1.20	1.10	1.17	2.69
2	1.53	1.16	1.20	2.80
3	1.52	1.11	1.17	2.73
4	1.53	1.13	1.14	2.77
5	1.24	1.13	1.14	2.78
6	1.22	1.14	1.17	2.77
7	1.22	1.17	1.18	2.79
8	1.22	1.15	1.14	2.78
Mean	1.34	1.14	1.16	2.76
Std. Dev.	0.16	0.024	0.020	0.036
3x Std. Dev.	0.48	0.072	0.060	0.12

B. OTHER QUALITY SYSTEMS FOR SELF EVALUATION

B.1 MEASUREMENT ASSURANCE PROGRAMS (MAPs)

NBS Measurement Assurance Programs (MAPs) are applications of documented precision, accuracy and statistical measurement control to measurement processes. MAP's statistically redundant procedures ascertain if individual measurements are "good enough" for their intended purpose [26]. At NBS, the error limit (or what is good enough) is determined by a balance between the cost of a better measurement and the possible economic loss from a less accurate measurement. "Good enough" means that the uncertainty of the measurement process is sufficiently small. When discussing the uncertainty of a measurement process, two characteristics apply: repeated measurements of the same quantity and the means of measurements by two different processes will disagree [27]. The process uncertainty will be a combination of the random variation in all the measurement parameters.

NBS developed MAPs for mass in 1963 and for length in 1972. These MAPs incorporate the basic methodology to establish process parameters, monitor the process control, update process parameters and maintain the calibration of National Standards. They make extensive use of check standards and a redundant measurement process monitored with control charts. The process control limits are established with historical data and the total uncertainty associated with the measurement process is computed with rigorous statistical methods.

The standards for length and mass are just two examples of National Standards maintained by NBS for reference use by industry, government and the scientific community [14]. Many activities require traceability to National Standards to insure measurements of "adequate accuracy." An example is the Department of Defense's Military Standard 45662 for procurement. Similar requirements have been established by the Nuclear Regulatory Commission, Food and Drug Administration, Environmental Protection Agency and National Aeronautics and Space Administration. These requirements have been established by balancing the cost of adequate measurements against the potential economic loss from less accurate measurements. Examples range from the analysis of ethylene dibromide (EDB) residues in food, with an incalculable cost in lives and health, to the possible measurement errors associated with the loss of a 75 million dollar communication satellite.

The NBS MAPs are a relatively costly means to establish precision, accuracy and measurement process control. MAPs may, therefore, not always be applicable to a broad-based materials testing laboratory where cost is a major concern. The application of statistical process control works well for clearly defined repetitive processes. A prime example is the analysis and plotting of data for check standards when analyzing steel production. The repetitiveness of the process reduces the individual analysis cost. However, a broad-based materials laboratory produces measurements that are both repetitive and infrequent. Using an analogy for manufacturing systems: process control for flexible short runs and truly random events, such as a fault in a silicon chip, are outside the statistical process quality control strategy [28]. The flexible short runs or infrequent measurements are typical of the measurements produced by a broad-based materials laboratory. The challenge, then, is to develop measurement assurance (precision, accuracy and process control) for these types of measurements.

B.2 LABORATORY ACCREDITATION

Laboratory accreditation is formal recognition that a testing laboratory is competent to carry out specific tests or types of tests on products or materials. Many U.S. and foreign accreditation systems have been developed, each with rules of procedure and management, to carry out laboratory accreditation (e.g., EPA, NIOSH, NVLAP). These accreditation systems have been developed because dependable test results facilitate commerce. Accreditation can ensure that tests are normally done correctly, and thereby increase the validity of test data, promote the acceptance of test data, and give technical credibility and status to a laboratory.

The competence of a laboratory is formally recognized only after all criteria of an accreditation system are met. The following accreditation criteria should be incorporated in a laboratory quality plan for self-evaluation. These are discussed in terms of their applicability to a self-evaluative quality plan for a broad-based materials laboratory. They include organization, personnel, facilities and environment, equipment and calibration, sample handling, methods and procedures, reports, records, and quality systems.

Organization Criteria

For in-house laboratories, criteria are usually developed, organized, formatted and distributed consistent with corporate policy. Thus, differences are expected for laboratories of different companies. Collecting and indexing this organizational information in compliance with the requirements of accreditation criteria would be an administrative burden with no benefit to laboratory quality. Therefore, organizational accreditation criteria are not recommended for a self-evaluative quality system.

Personnel Criteria

The three elements of personnel criteria that benefit a self-evaluative laboratory quality system are:

- . a formal technical training program to achieve uniform training of all personnel;
- . training in quality assurance procedures; and
- . ongoing documentation of all personnel training.

In addition to these formal systematic training programs, each new laboratory employee should be provided a quality orientation and all quality policies should be periodically reviewed with the entire staff. These personnel criteria can be easily adopted by a self-evaluative quality system.

Facility and Environmental Criteria

Facilities should be designed for the proper performance of the intended testing. The environmental conditions must be sufficiently controlled. The specific limitations (e.g., temperature, humidity, dust or fumes) should be described in test method or equipment instructions. Compliance with the facility and environmental criteria are applicable to a self-evaluative quality system. Additionally, housekeeping and employee safety are appropriate for laboratory policies, but the impact on a self-evaluative laboratory quality system is not well defined.

Equipment and Calibration Criteria

Without ongoing laboratory-wide programs for equipment maintenance and calibration, the validity of data generated within the laboratory cannot be assured. Equipment records should include equipment name, manufacturer's name and type identification, date received, date placed in service, current location and all maintenance details. Additionally, all measuring and test equipment must be calibrated before being placed in service and recalibrated with an established frequency. The calibration documentation should include procedures, facilities, measuring equipment, reference standards, date of last calibration, calibration reports, maximum time between calibration, date of next calibration and the traceability of standards. Reference standards should be used only for calibration. Other checks, appropriate to the test, should be performed during regular operation. Equipment and calibration records should be established and implemented as part of a self-evaluative laboratory quality system.

Sample Handling Criteria

Sample mix-ups or suspect identification can render high quality analyses meaningless. Therefore, a documented laboratory-wide sample handling procedure for sample traceability is essential to a self-evaluative laboratory quality system.

Test Method and Procedure Criteria

Accreditation criteria detailing requirements for documenting procedures, equipment operation, data recording, calculations and standards should be incorporated in a self-evaluative quality system as follows:

- . document all test methods, procedures, and operating procedures;
- . document laboratory equipment and calibrations;
- . record, process and report test results according to acceptable procedures;
- . eliminate errors associated with calculations, data transfers, and data processing techniques; and
- . control relevant standards.

Test Report Criteria

Laboratory reports should state that test results relate only to the items tested, and the measurement uncertainty (where relevant). The addition of the measurement uncertainties from MAEs is appropriate to a self-evaluative quality system. Additionally, a standardized report format would benefit the self-evaluation of laboratory quality.

Record Criteria

If record retention and retrieval systems, sufficiently detailed to ensure the ability to repeat any test, are corporate requirements for a broad-based materials laboratory, restatement in a quality plan would be a questionable duplication. However, any laboratory-specific exception should be included in the laboratory quality system.

Quality System Criteria

The operation of an ongoing quality assurance program effectively unites the individual accreditation criteria. A self-evaluative quality plan should, therefore, include a formal quality assurance program. Specific direction for documentation, accountability and review of the quality system are appropriate to self-evaluation. Fulfillment of these requirements by broad-based materials testing laboratory provides the framework for developing a self-evaluative quality system. Additional quality system criteria are:

- . proficiency testing;
- . procedures for handling client complaints;
- . procedures to remedy testing discrepancies; and
- . criteria for the evaluation and acceptance of subcontracted testing and calibration services.

B.3 INTERNAL QUALITY AUDIT

Central Laboratory's audit program was implemented in 1978. [9] The program, called "The Quality Audit," consists of sampling reports completed during the previous month and conducting a detailed evaluation of each report. Special emphasis is placed on equipment calibration, standards, test methods, sampling, test outcomes, report quality and record retention. The audit is conducted in a confidential, non-threatening manner between the auditor and contributing laboratory personnel. The findings are reviewed and agreed to by supervisors and selected laboratory engineers before they are forwarded to laboratory management. Laboratory management then responds to recommendations concerning quality system concerns.

Laboratory Organization

Central Laboratory is a large, non-research, broad-based materials testing laboratory. Its primary function is to support engineering, manufacturing and purchasing decisions about materials/components used in company products and manufacturing operations. Tested materials include metals, plastics, elastomers, paints, fuels, lubricants, textiles, adhesives, sealers, wastewater, and ceramics.

Section supervisors are members of a longstanding committee who discuss laboratory-wide problems and recommend corrective action or procedures. This committee plays a key role in the quality audit because changes to the audit procedure must have their approval. This contributes to supervisor commitment to audit recommendations and decisions since each and every supervisor is involved in these decisions. The three principal steps of the quality audit are: (1) evaluation and detection of discrepancies, (2) assessment, and (3) feedback and resolution of concerns.

Evaluation and Detection

The initial step in the quality audit is evaluation and detection of discrepancies. The auditor selects ten reports at random from approximately 700 reports per month. Each report is then evaluated to determine if the client's request has been satisfied. Then the standards, specifications, and test methods used are reviewed. After this background review, the auditor conducts in-depth reviews with the supervisor and contributing staff.

Notebooks and other section records are used to reconstruct the testing rationale. Special attention is given to sampling, testing, retesting, calibrations, standards, all calculations and record retention. Printouts from computerized equipment or plotters are reviewed. Sometimes it is necessary to retest to confirm the original data when the data indicates a possible error of misinterpretation. All potential discrepancies are identified during this phase of the audit.

Assessment

The second step in the quality audit is the assessment. All questionable results or practices are reviewed with the responsible supervisor and discrepancies are classified as significant or minor. A significant discrepancy is one which affects the conclusion of the report. All other discrepancies are considered minor.

Feedback - The Quality Audit Report

The auditor writes the quality audit report. This report includes statistical information on the discrepancy rate for the current audit and previous audits. These discrepancies are reported in general terms to preserve anonymity. The focus is to identify the discrepancy, not the individual staff. A balance between providing sufficient information to alert all laboratory staff to the type of discrepancy without specifically identifying the individual or section is maintained. The audit report also contains remarks which communicate recommendations revealed during the audit to laboratory management. The audit report, including auditor recommendations is reviewed and approved by the supervisor's committee before the audit report is distributed to laboratory management.

The auditor's recommendations are then discussed by the quality committee consisting of the department managers, the auditor and the chairman of the supervisors committee. The quality committee has the responsibility to resolve audit concerns and recommend changes as required to laboratory policy. Each recommendation the committee acts upon is reviewed extensively, with feedback from department supervisors and the supervisors committee, so that consensus decisions that benefit the entire laboratory are reached. This interactive discussion on quality audit recommendations has been a powerful tool to highlight report quality concerns.

Cost versus Benefit

It takes approximately 130 hours per month to conduct the quality audit, equivalent to about 1% of the cost of testing. This cost is small when compared to the continuing improvement in report quality, record retention, and quality procedures implemented as a result of the feedback and corrective actions generated. Because the quality audit is handled in a dynamic, interactive manner, it provides the framework to continually sample and improve the laboratory product, namely the laboratory report.

C. QUALITY SYSTEMS IN OTHER BROAD-BASED LABORATORIES

A number of corporate laboratories were visited to explore other techniques employed to assure quality. The specific programs used by those laboratories, applicable to a self-evaluative quality system, are qualitatively reviewed.

C.1 OIL/CHEMICAL INDUSTRY LABORATORY

This analytical services laboratory has a staff of more than 120 people divided into six analytical groups plus a supporting staff. The staff has 30 percent PhDs, 19 percent MS degrees, 33 percent BS degrees, and 18 percent non-degreed technicians. The hiring trend has been to increase the percentage of PhDs. A computer services staff of about 60 provide support to the testing operations. The laboratory budget totaled about \$125 million, with 5% devoted to capital equipment acquisitions. Quality assurance emphasis was observed in two of the six analytical groups: environmental and petroleum. The additional staff responsible for the quality assurance system included a procedure writer, a statistician, and a representative from the computer service staff. The computer services included were sample tracking and management information.

Both the environmental and petroleum service groups employed auditing procedures. The petroleum group utilized two types of audits: (1) a quality assurance audit, which retested samples from various quality control laboratories; and (2) a method audit, which checks data precision by conducting round-robin analyses with six quality control laboratories. The statistician compiled data and management reports for these audits. The environmental group used round-robin analyses with the quality control laboratories until about 5 years ago. At that time a mobile laboratory audit was initiated to provide on-site training in addition to the audit.

The following are related observations which might affect the overall laboratory quality:

- . the laboratory director maintained technical awareness by attending relevant technical conferences;
- . A dual career ladder was provided;
- . good personnel, quality assurance, the method audit and participation in ASTM and quality circles were cited as keys to quality; and
- . a quality assurance manual, based on ASTM E-548, was used.

This laboratory has successfully implemented a quality system in two analytical groups: environmental and petroleum. The environmental group was able to justify the cost of the quality system by reducing the company's liability for out-of-compliance EPA fines. In contrast, the petroleum group economically justified the quality program with the close control of expensive product ingredients at several manufacturing locations. These two groups, however, had the only systematic quality program observed. The balance of the laboratory relied on the "technical expert" approach to laboratory quality.

C.2 BUILDING MATERIALS LABORATORY

This building materials laboratory included a staff of 45, divided into three groups. The staff's education was evenly divided among PhDs, BS degrees, and

non-degreed technicians. The trend was to hire more PhDs. The analytical services budget totaled about \$4 million, with 15% devoted to capital equipment acquisitions. The quality assurance effort was focused on the analysts. Of the staff with quality system responsibility, about 5% are concerned with quality assurance, 15% with method development, and about 20% with a monthly audit of three representative reports. This monthly report audit, however, was initiated a few years ago and appeared to be rapidly evolving. The evolving quality system included policy-making, quality assurance manual documentation, operating procedures, method documentation, audits, and a NVLAP accreditation.

The quality assurance was initiated by a new manager based on a strategic analysis of laboratory business. Twelve visits to other company laboratories were utilized to identify quality ideas for implementation. This evaluation identified the need for a quality system. The key elements identified are:

- . participation in the American Laboratory Manager Association (ALMA);
- . accreditation by NVLAP with the help of a consultant;
- . adapting an internal report quality audit from that described in appendix B.3;
- . improved client communication improved with: (1) feedback on audited reports, (2) monthly reprioritized jobs, and (3) blind samples as proficiency tests;
- . an "Open Lab" policy allowed clients to perform analyses after proper training;
- . accomplishments were reported weekly to upper management;
- . a dual career ladder was provided; and
- . two performance review systems were used.

While this building materials laboratory has implemented important parts of a self-evaluative quality system (i.e., the audit, accreditation and an increased awareness of quality concerns), a comprehensive all-inclusive quality system was not observed.

C.3 CHEMICAL INDUSTRY LABORATORY

This laboratory utilized the efforts of 165 people divided into six groups. The staff had 30 percent PhDs, 20 percent MS degrees, 50 percent BS degrees, plus a few non-degreed technicians. The trend was to hire more PhDs. The capital budget was \$1-2 million. The quality assurance efforts originated from corporate staff and had been in operation over 25 years. The responsibility for quality assurance was documented in a corporate manual, "Policy and Program for Quality Assurance". This manual also documented the responsibility of and the reporting relationships for marketing, technical service, development, purchasing, manufacturing, quality control, distribution, and the executive department. For example, within the technical service, the analytical services laboratory managers were responsible for quality assurance. This responsibility was supported by corporate policies and procedures documented in an "Analytical Laboratory Operations Manual".

Some of the policies and procedures cited are:

- . verbal communication between the analyst and the client was required before any work is accepted;
- . the quality assurance program included : (1) validation of sampling and analysis, (2) standardization, (3) calibration, and (4) documentation;
- . various audits were conducted with different frequencies.
- . shared product specification development progressed from developmental specifications during field tests to product specifications and finally to sales specifications (all specifications required validated methods with documented precision statements); and
- . employee quality circles and statistical process control procedures were encouraged by product area managers.

The comprehensive documentation on laboratory procedures and operations in the "Analytical Laboratory Operation Manual" included sections analogous to the ASTM E-548 criteria with detailed specific categories (i.e., analytical standards, compilation of spectra, nomenclature, notebook keeping, purchasing, etc.). This laboratory had an extensive quality system permeating research, development and product testing laboratories.

C.4. SUBCONTRACT LABORATORIES

While conducting research on laboratory quality systems, the need arose to purchase water analyses. This necessitated the evaluation of potential laboratories. The laboratory evaluation objectives were: (1) to assure the laboratory's ability to produce accurate results in compliance with EPA methods; (2) to identify turnaround time for results; and (3) to justify the chosen laboratory to purchasing if it was not the low bidder. The evaluation was limited to half day visits at local contract laboratories. The evaluation schedule, the measurement quality concerns, and the importance of analysis turnaround time were discussed with purchasing before the evaluation began.

The evaluation criteria were developed by technical laboratory people, combining ASTM E-548 criteria and experience in conducting three-day supplier quality assurance audits. A numerical rating system, with 50 points for a system audit based on ASTM E-548 criteria and 50 points for proficiency testing was developed. Some detail for this rating included the following specific items:

- . the laboratory's ability judged against ASTM E-548 criteria;
- . the measurement ability based on documentation of internal or external performance tests (e.g., round-robins, EPA quarterly samples, or participation in ASTM); and
- . the laboratory competence judged from management commitment to quality policies, recognition systems for technical staff and fair, accountable follow-up on administrative or technical procedures.

The assignment of a numerical rating was subjective, but it was balanced by averaging the numerical scores of all audit team members. Each team member

asked questions and verified different items under each category. The numerical results are summarized in exhibit C.4. The numerical rating of the contract labs provide the basis for ranking the labs. The top two contract labs were then reviewed for potential risks. Purchasing, laboratory management and technical laboratory personnel then discussed the evaluation of the potential subcontract laboratories. A consensus decision was made to split the work between two laboratories.

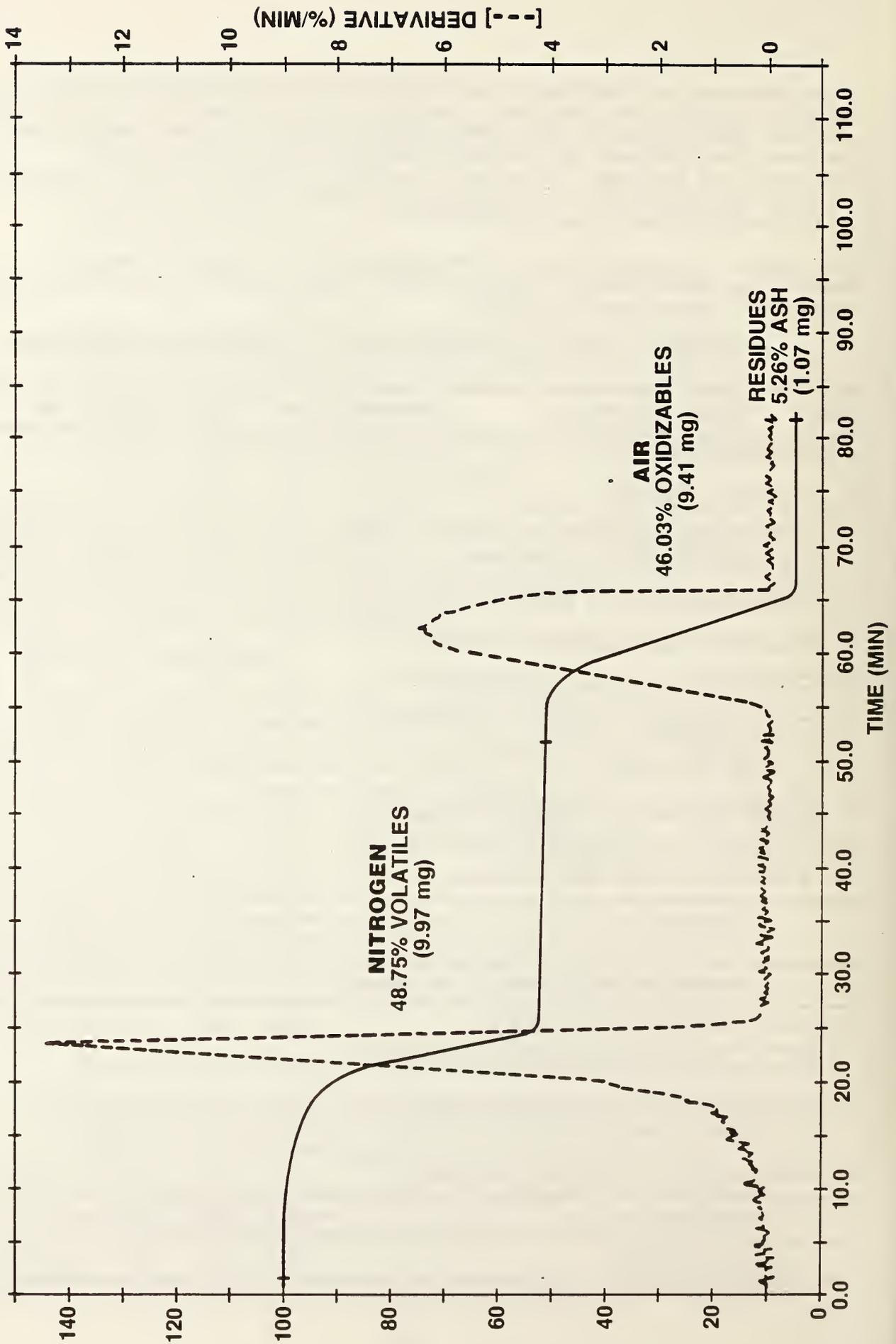
The judgment of acceptability of the contract laboratories was made by considering three criteria: accreditation, proficiency testing and laboratory management's commitment to quality.

Exhibit C.4 - QUALITY ASSURANCE AUDIT - WATER TESTING LABORATORIES

Part I - System - 50 Points	Laboratory-> #1	#2	#3	#4	#5
1. Equipment (5 points)	4.3	4.7	4.3	4.0	3.0
2. Personnel (5 points)	5.0	5.0	5.0	5.0	4.5
3. Calibrations-pipets /instruments/balances (5 points)	4.7	3.7	1.5	3.3	3.7
4. Standards/Reagents (5 points)	2.7	3.0	1.6	3.0	4.0
5. Cleanliness Lab/Glassware (5 points)	1.7	3.0	3.0	3.0	4.5
6. Sample-Chain of Custody (5 points)	4.7	5.0	3.0	4.0	3.0
7. Record Keeping (5 points)	5.0	4.7	3.6	3.2	4.0
8. QC Manual (5 points)	4.3	4.7	4.0	3.0	0.0
9. Reporting Capability-Computers (5 points)	4.3	4.0	4.3	2.0	4.0
10. Procedures-Documentation & verification of use (5 points)	3.7	3.5	1.6	2.3	2.5
SUBTOTALS	40.4	41.3	31.9	32.8	33.2
Part II - Capability Verification - 50 points					
1. EPA QC Performance (25 points) Trace Metals WPI-three levels (15 points) EP Metals-two levels (10 points)	18.0	14.0	13.0	7.0	3.0
2. QC Performance - Internal (15 points) Duplicates/Spikes/Blanks	13.0	9.0	10.0	10.0	9.0
3. QC Performance - External (10 points)	7.3	0.0	8.0	8.0	0.0
SUBTOTAL	38.3	23.0	31.0	25.0	12.0
GRAND TOTALS	78.7	64.3	62.9	57.8	45.2

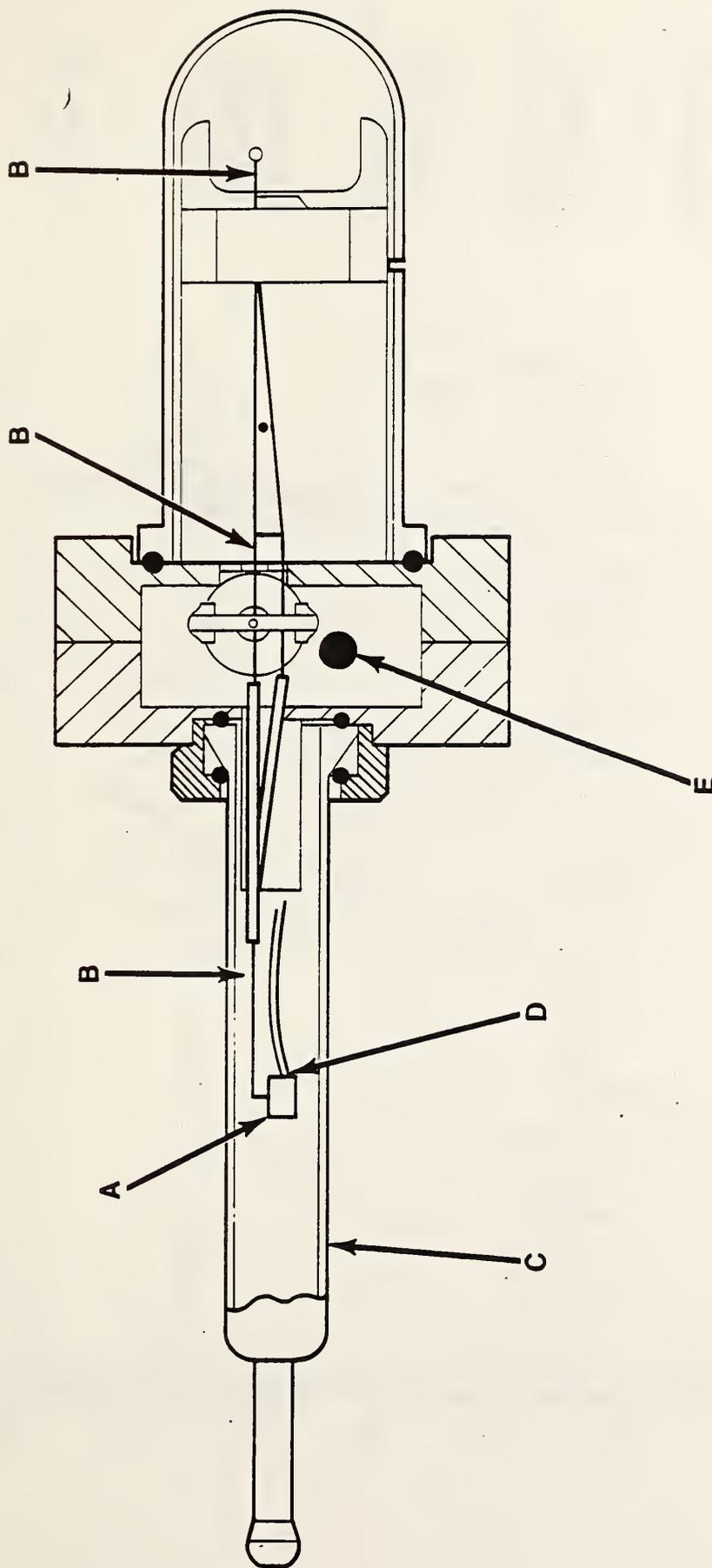
THERMOGRAM (WEIGHT LOSS PROFILE)

Figure 1



DUPONT 951 THERMOGRAVIMETRIC ANALYZER BALANCE ASSEMBLY

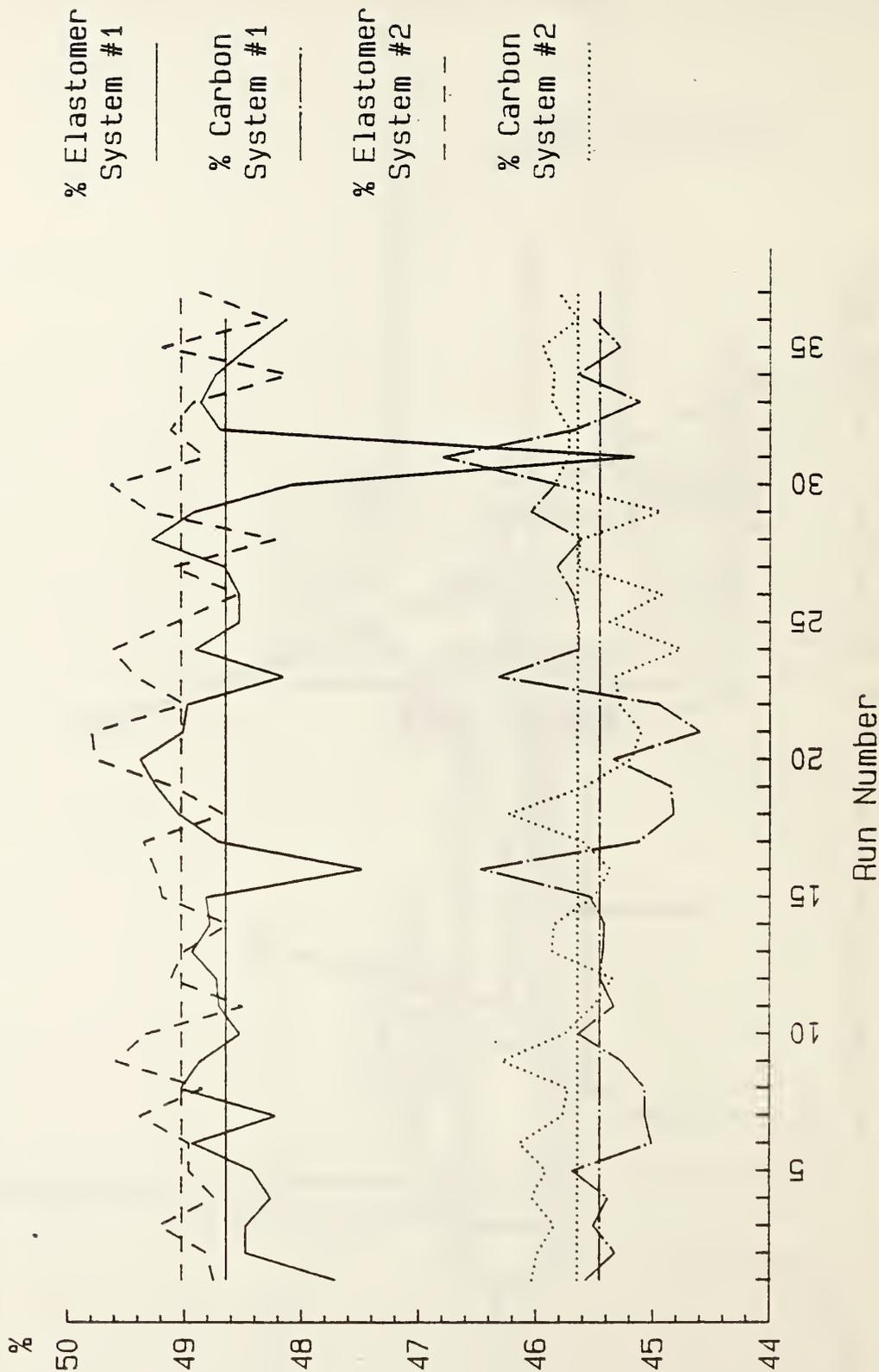
Figure 2



- A. PLATINUM SAMPLE PAN
- B. BALANCE ARM
- C. QUARTZ FURNACE TUBE
- D. SAMPLE THERMOCOUPLE
- E. PURGE GAS INLET

Figure 3

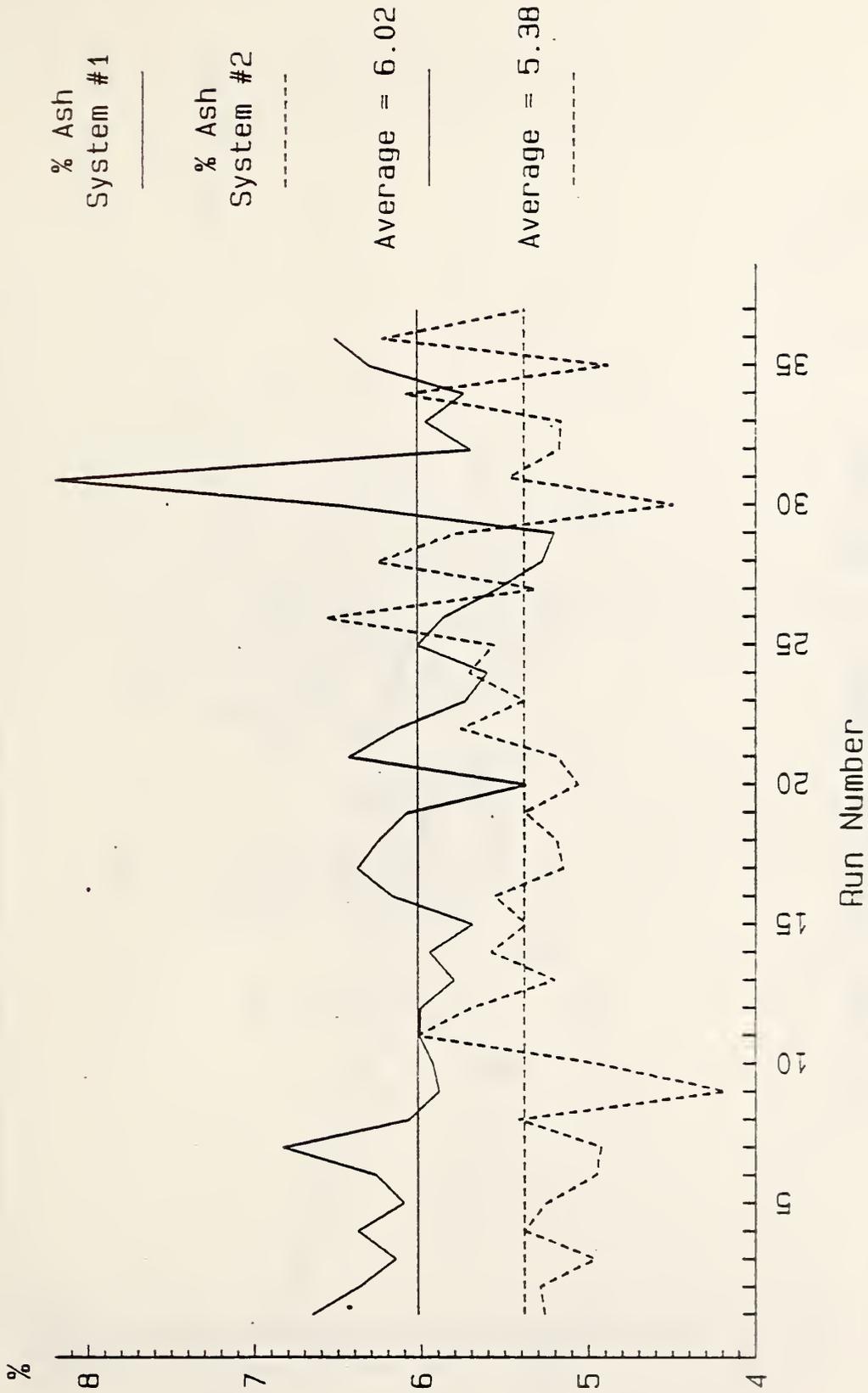
Run Chart for TGA Systems 1 and 2 % Elastomer and % Carbon



System 1 - 36 runs
System 2 - 37 runs

Figure 4

Run Chart for TGA Systems 1 & 2 % Ash Content



System 1 - 36 runs
System 2 - 37 runs

Figure 5

% Ash versus Sample Weight

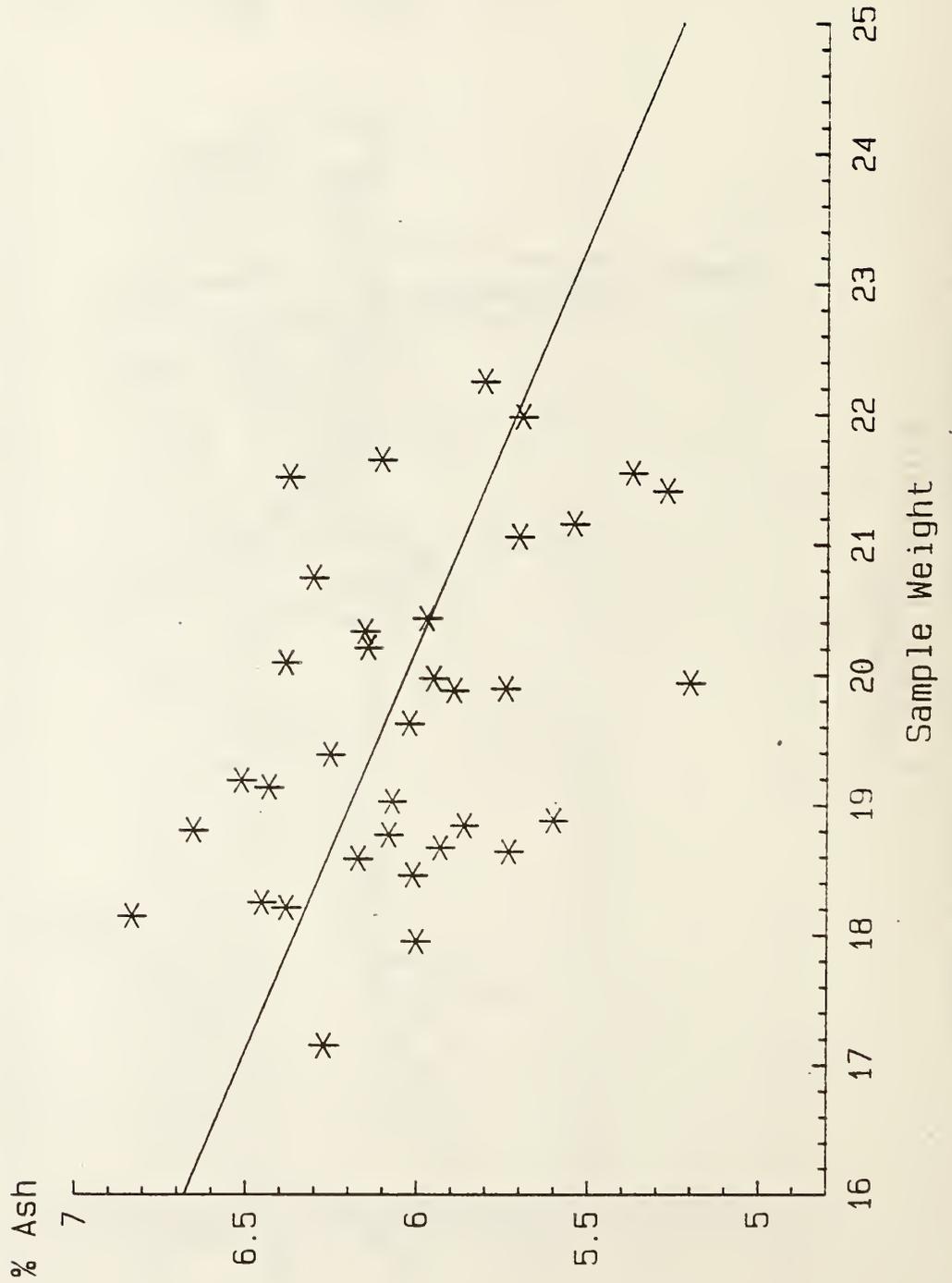


Figure 6

% Carbon versus % Elastomer TGA Systems 1 and 2

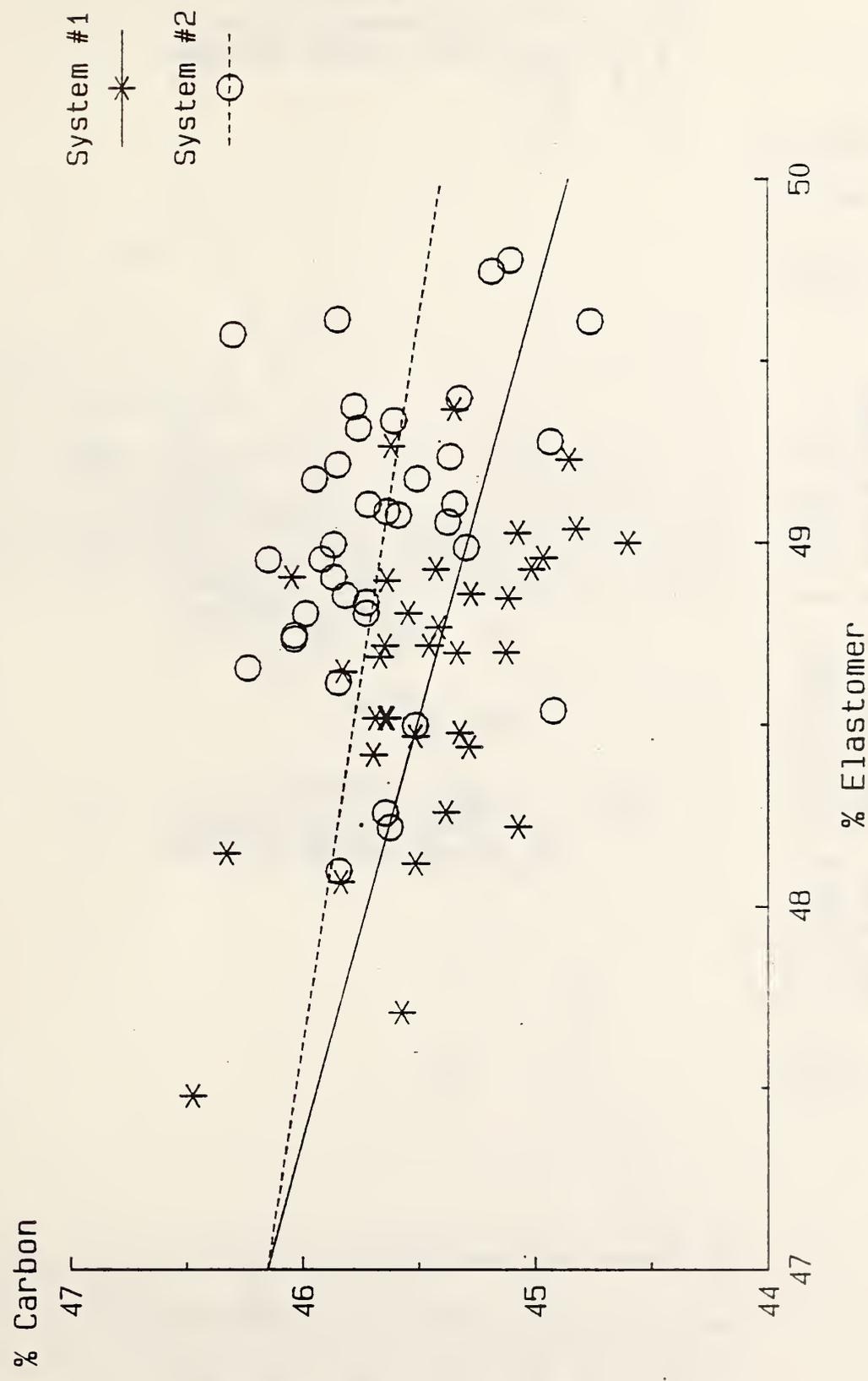


Figure 7

% Ash versus % Elastomer TGA Systems 1 and 2



Figure 8
Quality Comparison
English Master Set 1

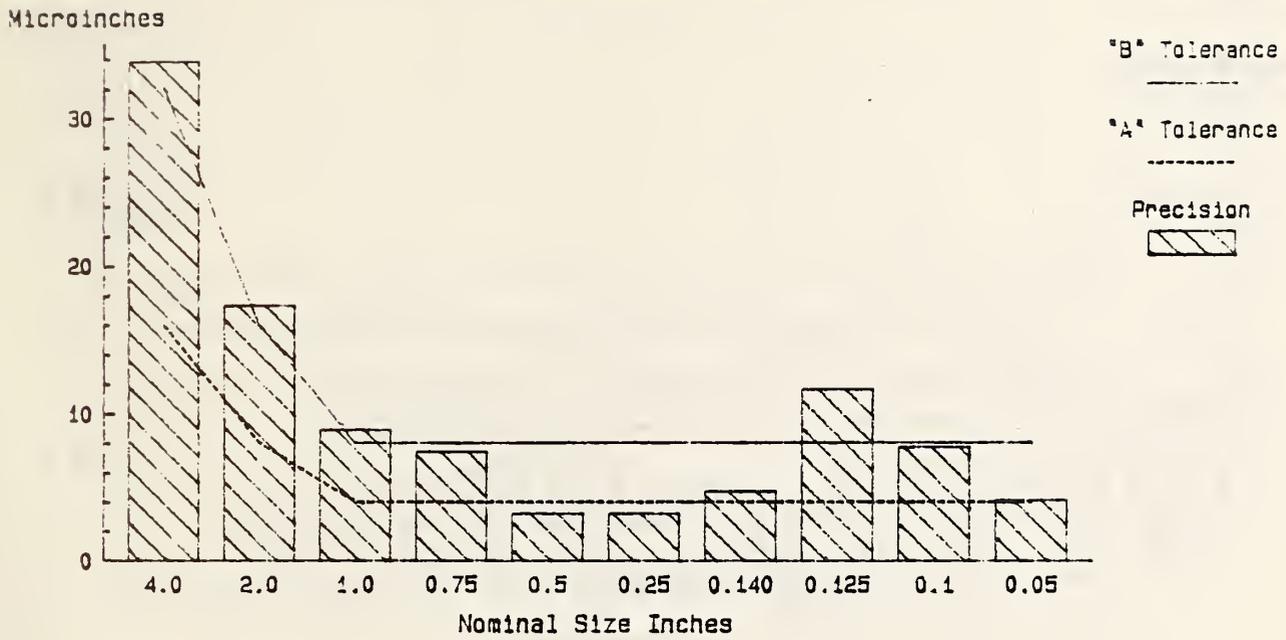


Figure 9
Quality Comparison
English Master Set 2

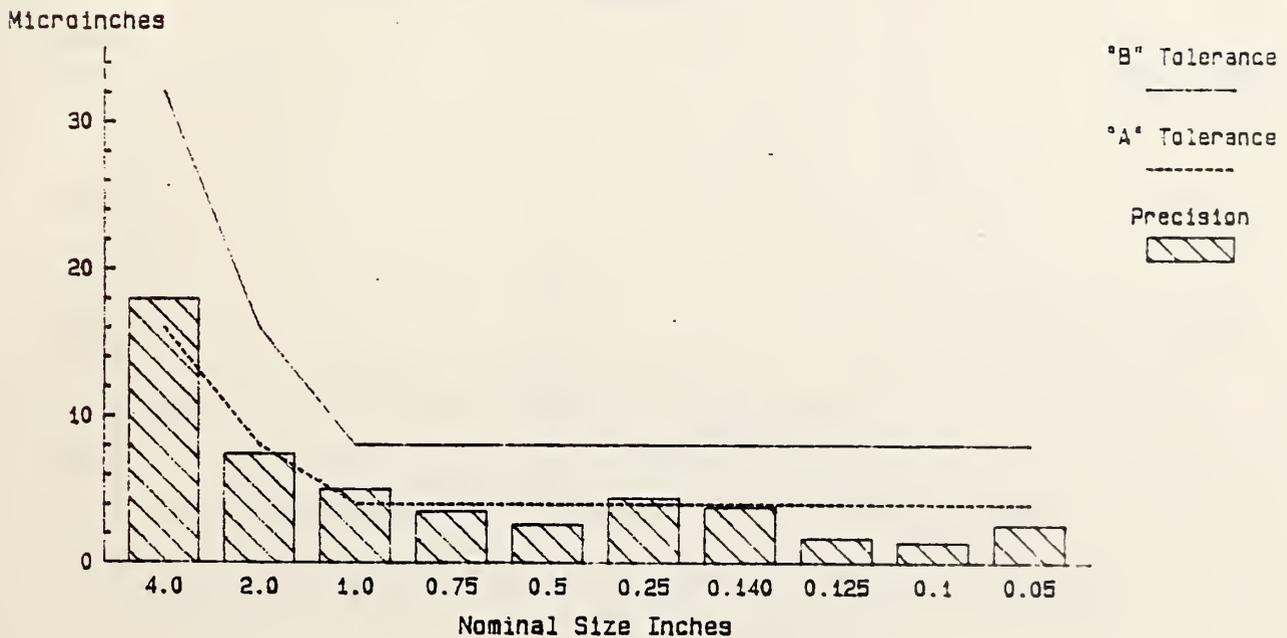


Figure 10
 Quality Comparison
 Metric Master

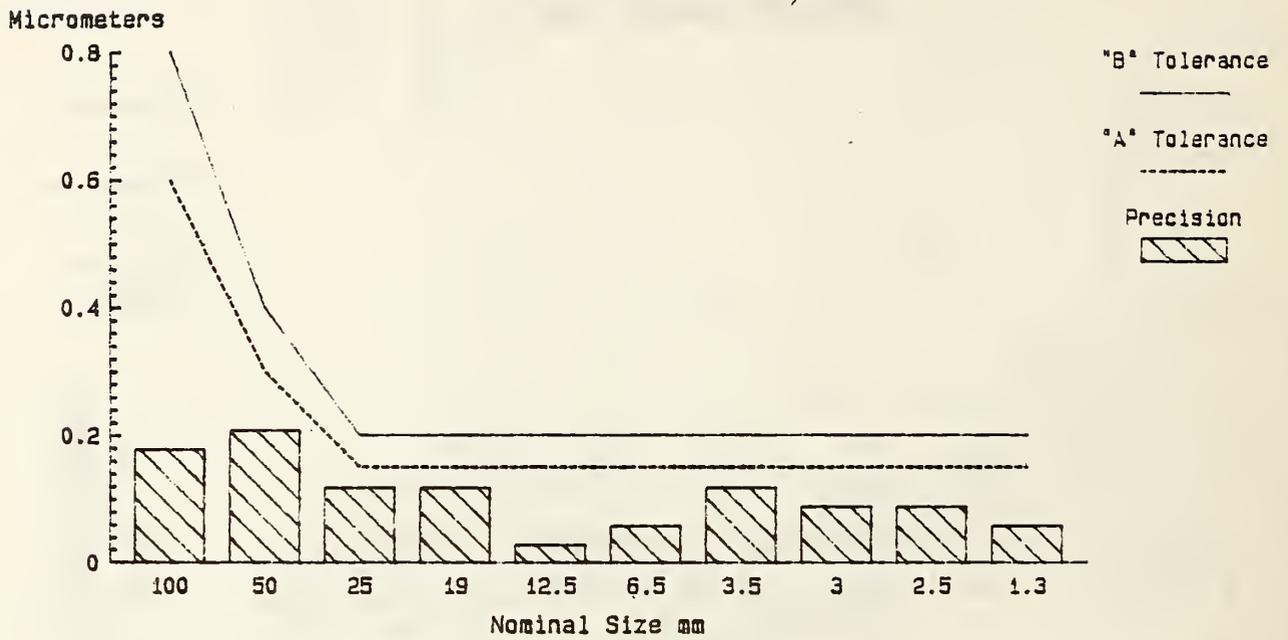


Figure 11

LECO Carbon Analysis Control Chart
Cast Iron NBS 3c - 2.3% +/- 0.014

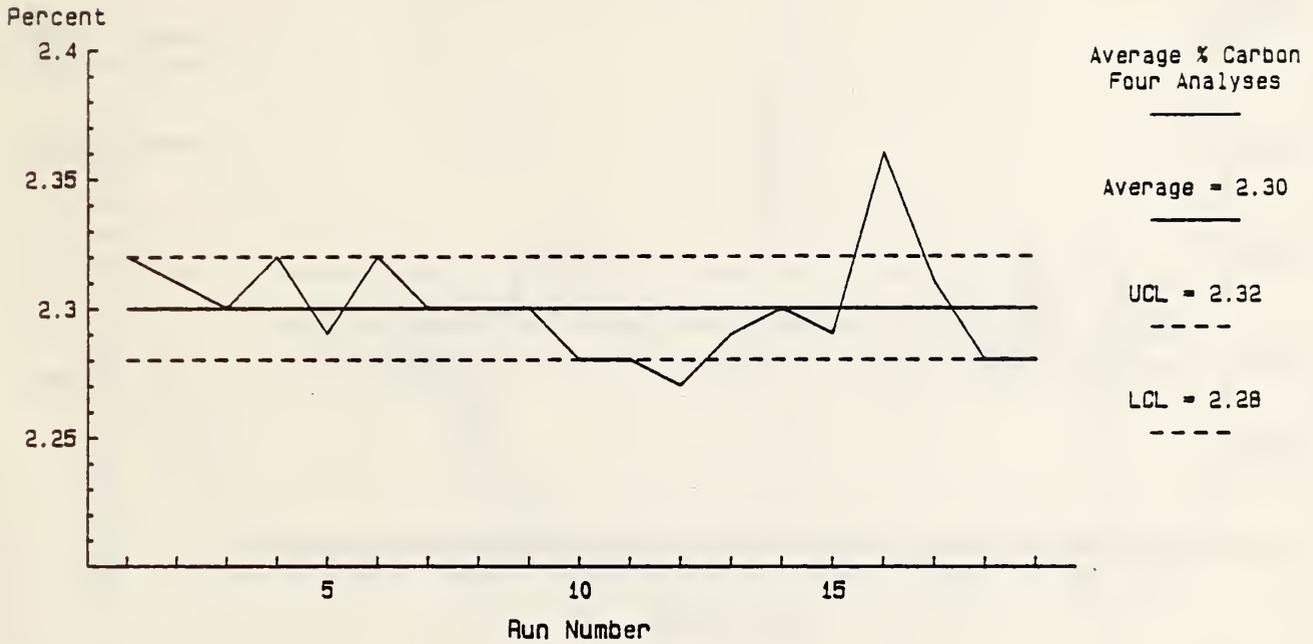


Figure 12

LECO Carbon Analysis Control Chart
Cast Iron NBS 3C - 2.3% +/- 0.014
Range

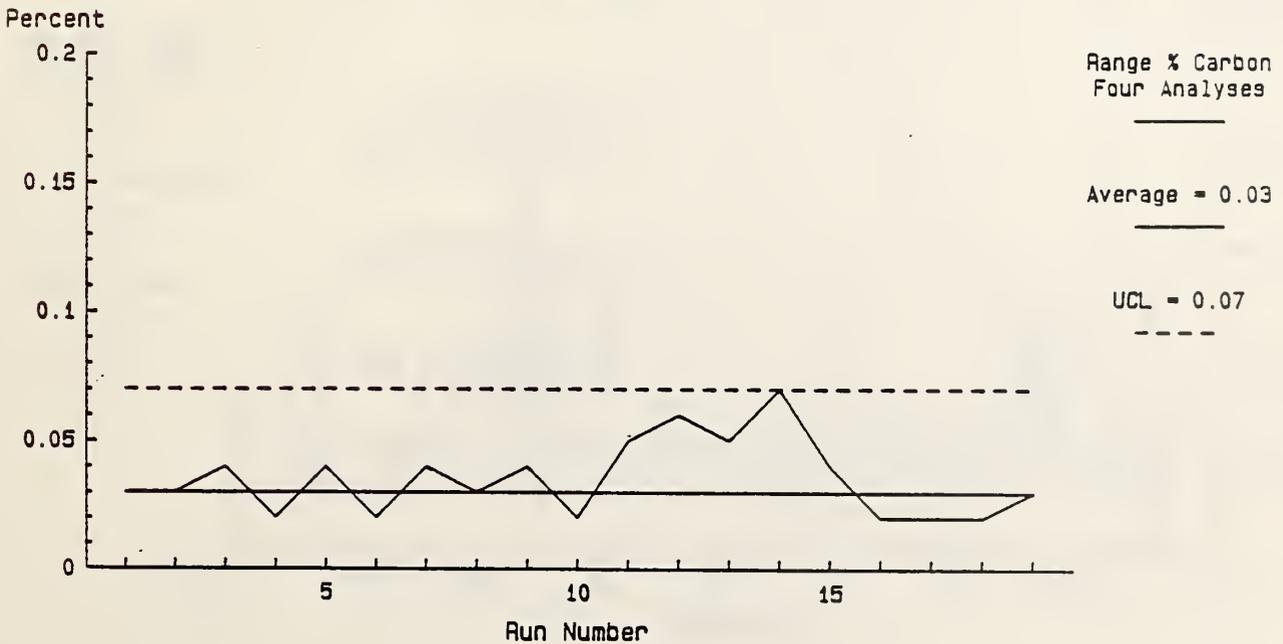


Figure 13

LECO Carbon Analysis Control Chart
Cast Iron Ar 302 - 3.30 % +/- 0.03
Averages

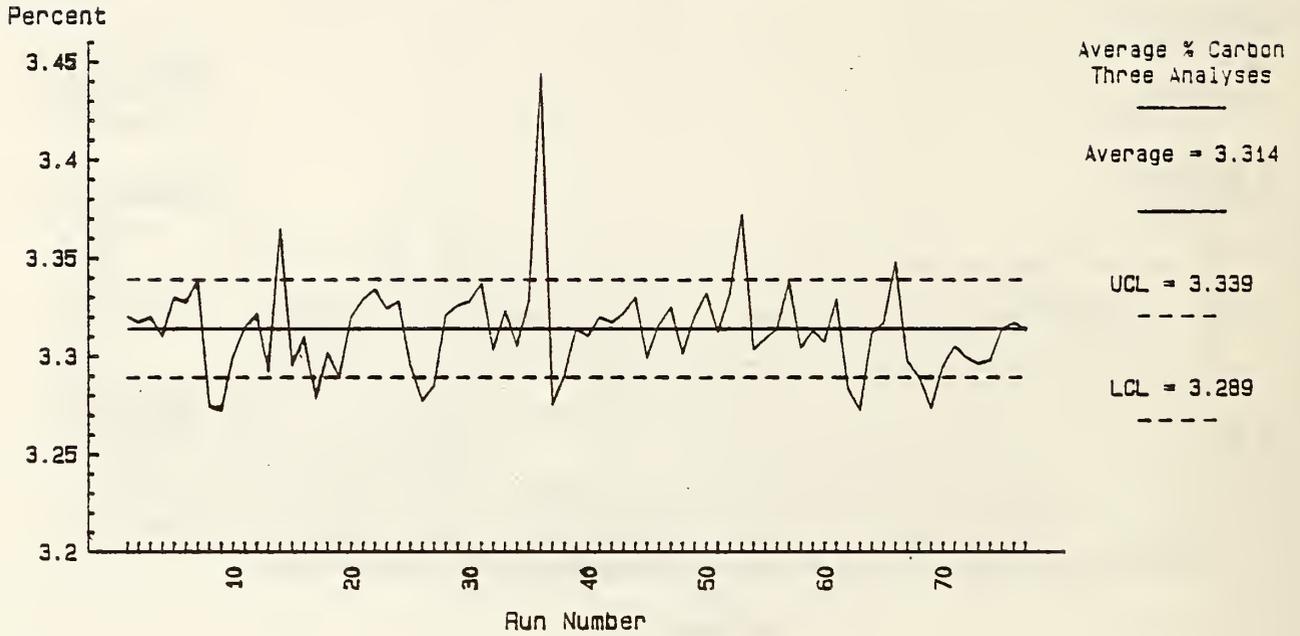


Figure 14

LECO Carbon Analysis Control Chart
Cast Iron Ar 302 - 3.30% +/- 0.03
Range

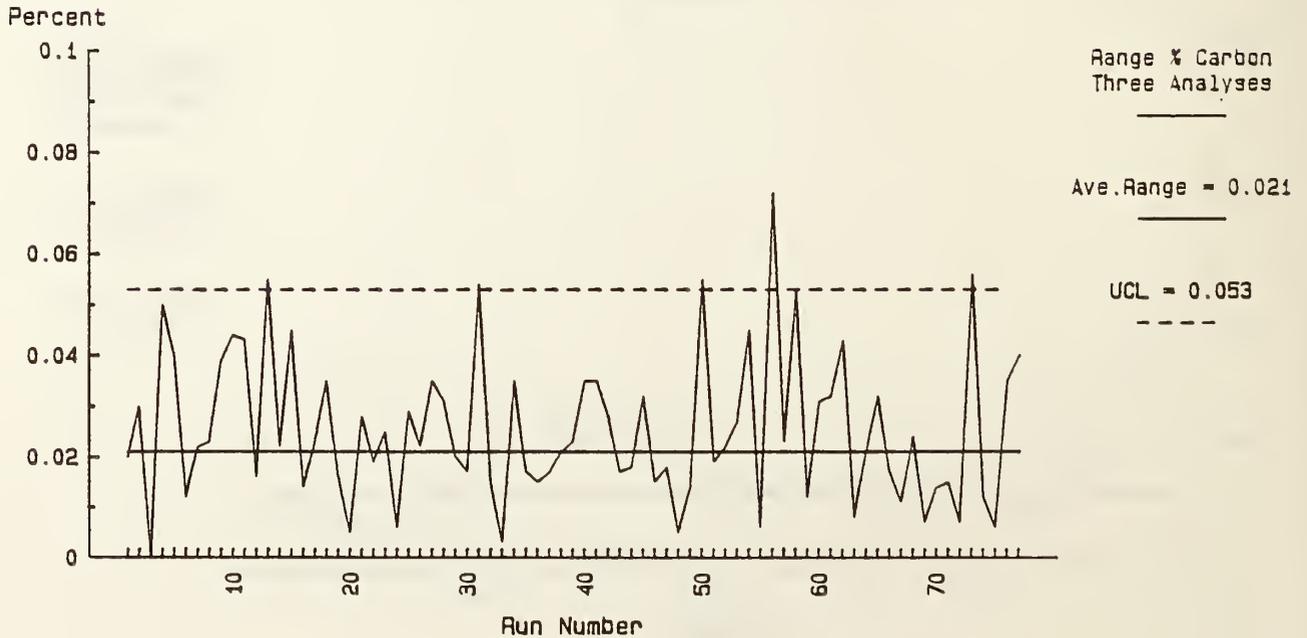
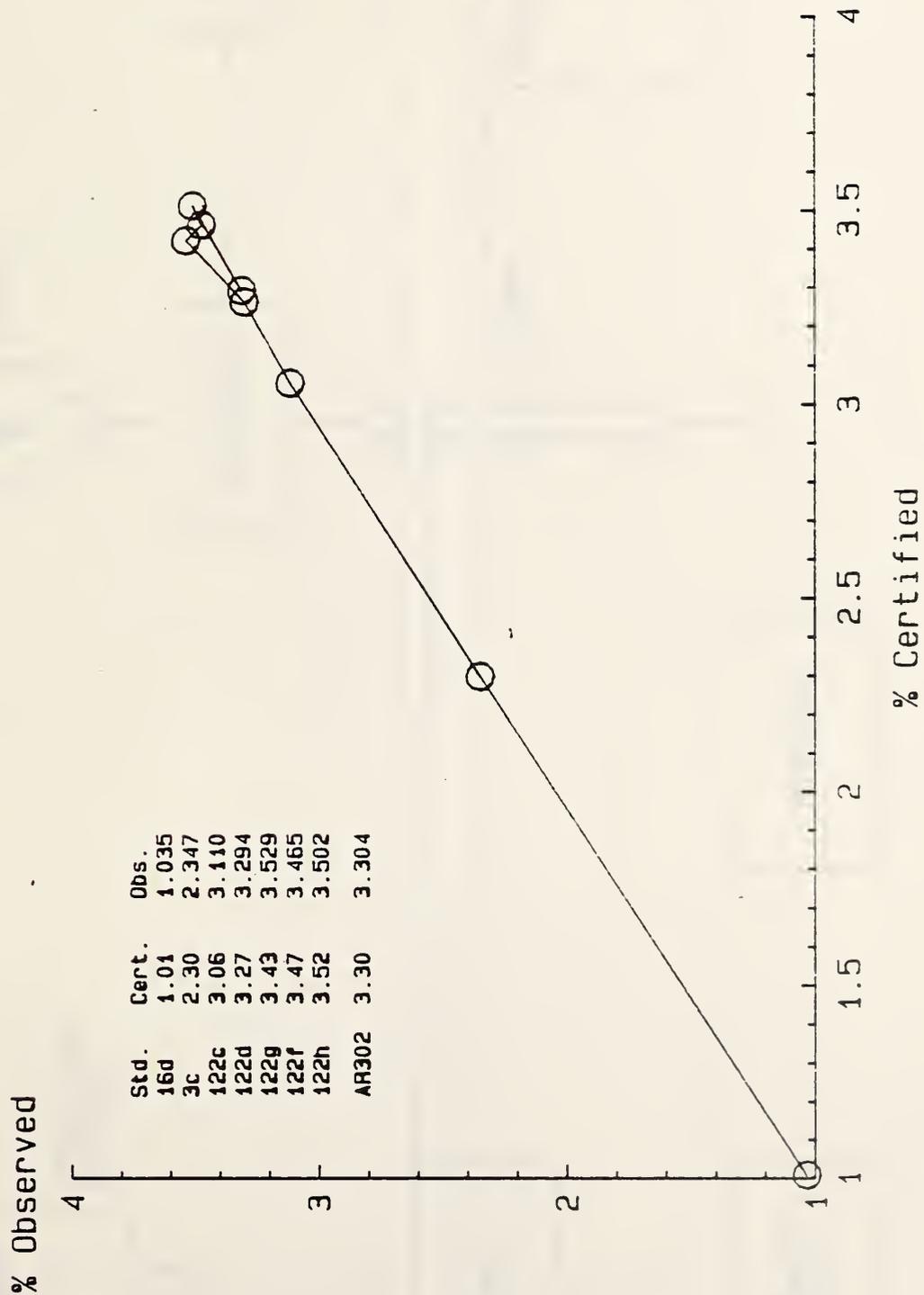


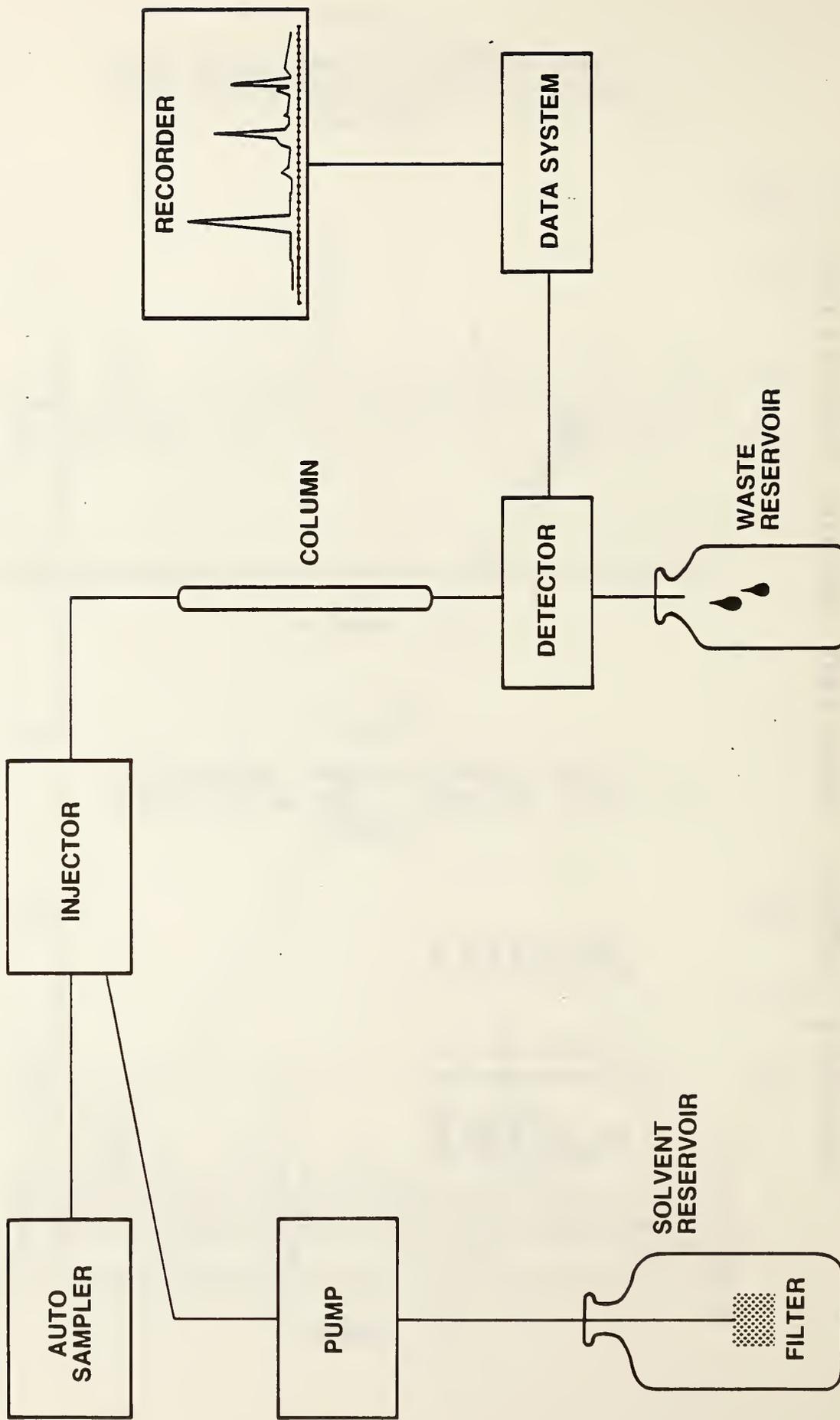
Figure 15

Percent Carbon Observed versus Certified



SCHEMATIC: HPLC INSTRUMENTATION

Figure 16



TYPICAL CHROMATOGRAM ENGINE COOLANT WITH BOTH TT AND MBT

Figure 17

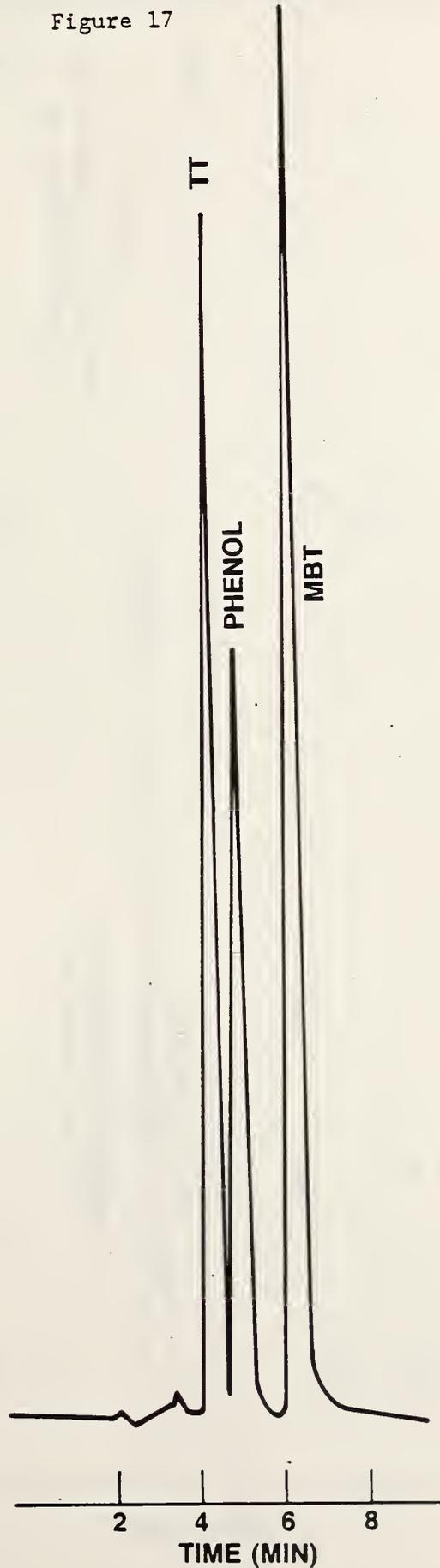


Figure 18

Figure 18
**Variability of Measured TT
 Union Carbide Reference Standard**

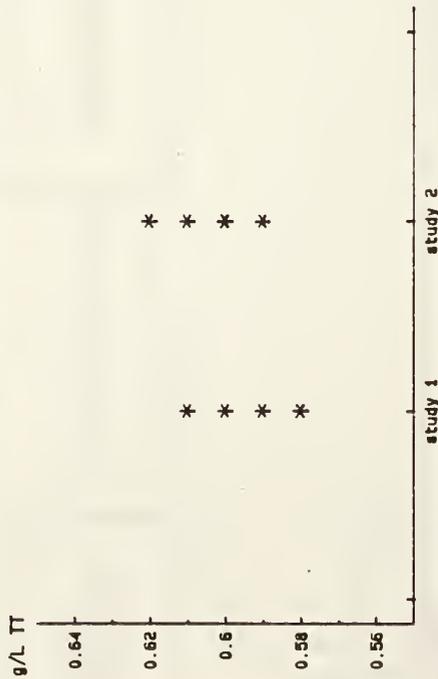


Figure 19

Figure 19
**Variability of Measured TT
 Quarterly Sample #1**

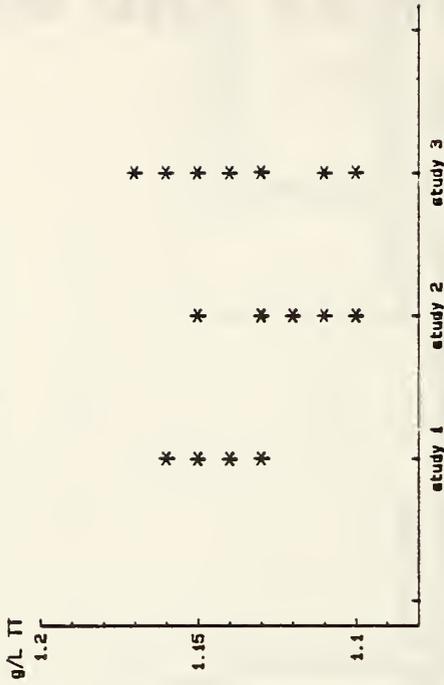


Figure 20

Figure 20
**Variability of Measured TT
 Quarterly Sample #2**

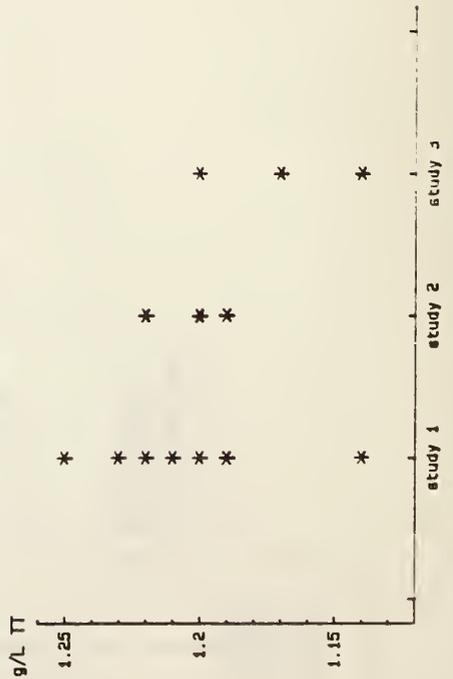
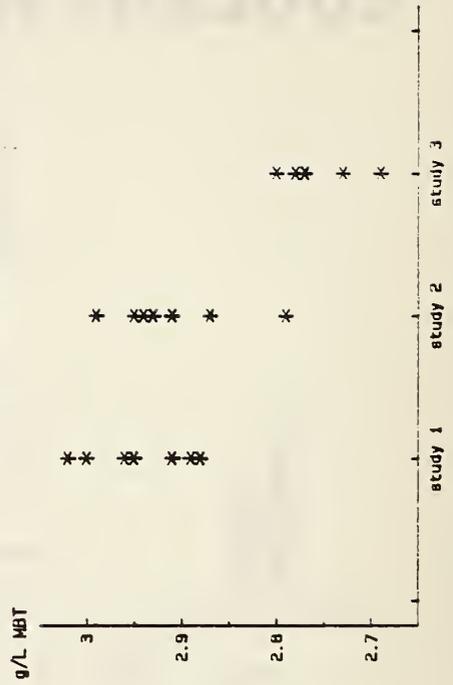


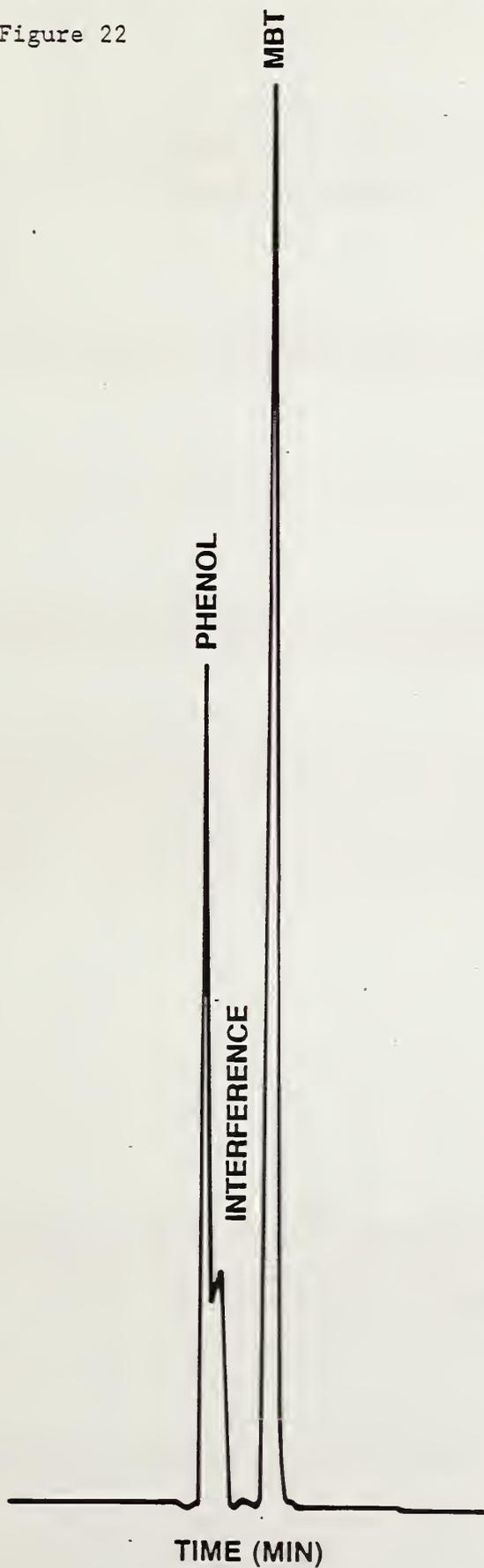
Figure 21

Figure 21
**Variability of Measured MBT
 Quarterly Sample #3**



MBT STANDARD REFERENCE WITH PHENOL INTERFERENCE

Figure 22



U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET <i>(See instructions)</i>	1. PUBLICATION OR REPORT NO. NBSIR-85/3278	2. Performing Organ. Report No.	3. Publication Date November 1985
4. TITLE AND SUBTITLE A Self-Evaluative Laboratory Quality System			
5. AUTHOR(S) C.J. Kelly; K.D. Bruley; D.H. Craig; D.J. Pangonis; J.W. Locke			
6. PERFORMING ORGANIZATION <i>(If joint or other than NBS, see instructions)</i> NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20134 Gaithersburg, MD 20899		7. Contract/Grant No.	8. Type of Report & Period Covered
9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS <i>(Street, City, State, ZIP)</i>			
10. SUPPLEMENTARY NOTES <input type="checkbox"/> Document describes a computer program; SF-185, FIPS Software Summary, is attached.			
11. ABSTRACT <i>(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)</i> This report describes the evaluation of Measurement Assurance Experiments (MAEs) for determining the quality of within-laboratory test data. A general self-evaluative quality system is outlined and objective measures of data quality, precision and/or accuracy are presented for four fully described MAEs. Measurement Assurance Programs (MAPs), laboratory accreditation, and internal quality audits are also discussed.			
12. KEY WORDS <i>(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)</i> laboratory accreditation; measurement assurance; quality; quality assurance; quality audit; quality control; system quality			
13. AVAILABILITY <input checked="" type="checkbox"/> Unlimited <input type="checkbox"/> For Official Distribution. Do Not Release to NTIS <input type="checkbox"/> Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. <input checked="" type="checkbox"/> Order From National Technical Information Service (NTIS), Springfield, VA. 22161		14. NO. OF PRINTED PAGES 63	15. Price \$9.95

